

# A Beta Way: A Tutorial For Using Beta Regression in Psychological Research

Jason Geller<sup>1</sup>, Robert Kubinec<sup>2</sup>, Chelsea M. Parlett Pelleriti<sup>3</sup>, and Matti Vuorre<sup>4</sup>

<sup>1</sup>Department of Psychology and Neuroscience, Boston College

<sup>2</sup>University of South Carolina

<sup>3</sup>Canva

<sup>4</sup>Tilburg University

## Abstract

Rates, percentages, and proportions are common outcomes in psychology and the social sciences. These outcomes are often analyzed using models that assume normality, but this practice overlooks important features of the data, such as their natural bounds at 0 and 1. As a result, estimates can become distorted. In contrast, treating such outcomes as beta-distributed respects these limits and can yield more accurate estimates. Despite these advantages, the use of beta models in applied research remains limited. Our goal is to provide researchers with practical guidance for adopting beta regression models, illustrated with an example drawn from the psychological literature. We begin by introducing the beta distribution and beta regression, emphasizing key components and assumptions. Next, using data from a learning and memory study, we demonstrate how to fit a beta regression model in R with the Bayesian package `brms` and how to interpret results on the response scale. We also discuss model extensions, including zero-inflated, zero- and one-inflated, and ordered beta models. Basic familiarity with regression modeling and R is assumed. To promote wider adoption of these methods, we provide detailed code and materials at [https://github.com/jgeller112/beta\\_regression\\_tutorial](https://github.com/jgeller112/beta_regression_tutorial).

**Keywords:** beta regression, beta distribution, R tutorial, Psychology, Learning and memory

Jason Geller  <https://orcid.org/0000-0002-7459-4505>

Robert Kubinec  <https://orcid.org/0000-0001-6655-4119>

Chelsea M. Parlett Pelleriti  <https://orcid.org/0000-0001-9301-1398>

Matti Vuorre  <https://orcid.org/0000-0001-5052-066X>

The analyses herein were not preregistered. Data, code, and materials for this manuscript can be found at [https://github.com/jgeller112/beta\\_regression\\_tutorial](https://github.com/jgeller112/beta_regression_tutorial). The authors have no conflicts of interest to disclose. Author roles were classified using the Contributor Role Taxonomy (CRediT; <https://credit.niso.org/>) as follows: Jason Geller: Conceptualization, Data curation, Formal analysis, Project administration, Resources, Visualization, Writing - original draft; Robert Kubinec: Formal analysis, Validation, Writing - review & editing; Chelsea M. Parlett Pelleriti: Formal analysis, Writing - review & editing; Matti Vuorre: Formal analysis, Resources, Supervision, Validation, Writing - review & editing

Correspondence concerning this article should be addressed to Jason Geller, Department of Psychology and Neuroscience, Boston College, McGuinn 300z, Chestnut Hill, MA 2467, USA, Email: [drjasongeller@gmail.com](mailto:drjasongeller@gmail.com)

 Preprint

This manuscript is currently **under review** and has not been peer-reviewed. Content is **subject to change**. Please feel free to provide feedback!

## Introduction

Many outcomes in psychological research are naturally expressed as proportions or percentages. These include the proportion of correct responses on a test (e.g., Kornell and Bjork (2008)), the proportion of time a participant fixates on a stimulus in an eye-tracking task (e.g., James et al., 2025), or the proportion of respondents endorsing a particular belief (e.g., Costello et al., 2024). A central question for studies involving proportional data is how to analyze these outcomes in a statistically appropriate and interpretable way.

Researchers frequently default to linear models that assume Gaussian (normal) distributions, such as t-tests, ANOVAs, and linear regression. However, these models make strong assumptions: (1) residuals are normally distributed, (2) the outcome is unbounded (from  $-\infty$  to  $\infty$ ), and (3) variance is constant across the range of the data. These assumptions are rarely satisfied in practice (Sladekova & Field, 2024), and they are especially ill-suited for proportional outcomes, which are bounded between 0 and 1 and often exhibit heteroscedasticity—non-constant variance, particularly near the boundaries (Ferrari & Cribari-Neto, 2004; Paolino, 2001; Smithson & Verkuilen, 2006). Violating these assumptions can lead to biased estimates and spurious inferences, especially when many observations cluster near 0 or 1.

In some cases, a generalized linear model (GLM) can relax the assumption of normality. For example, binomial and Bernoulli models—often referred to as logistic regression when used with a logit link—are appropriate when the outcome is binary (0/1) or represents the number of successes out of a fixed number of trials. These models, however, rely on data that arise from discrete trial structures. They tend to perform poorly when the observed proportions are truly continuous or when the data show extra variability (overdispersion), particularly when many values occur near the boundaries of the scale (0 and 1).

The challenges of analyzing proportional data are not new (see Bartlett, 1936). Fortunately, several existing approaches address the limitations of commonly used models. One such approach is beta regression, an extension of the generalized linear model that employs the beta distribution (Ferrari & Cribari-Neto, 2004; Paolino, 2001). Beta regression offers a flexible and robust solution for modeling proportional data directly by accounting for boundary effects and over-dispersion, making it a valuable alternative to traditional binomial models. This approach is particularly well-suited for psychological research because it can handle both the bounded nature of proportional data and the non-constant variance often encountered in these datasets (Sladekova & Field, 2024). In addition, the direct modeling of proportions allows comparability across tasks and scales, and can be particularly valuable when only the proportional data is available, as is often the case with secondary data that lack item-level structure or point values.

While in this paper we will focus on proportional-responses that lie between 0 and 1—it is important to note that our analysis applies to any bounded continuous scale. Any bounded scale can be mapped to lie within 0 and 1 without resulting in a loss of information as the transformation is linear.<sup>1</sup> Consequently, a scale that has natural end points of -1,234 and +8,451—or any other end points on the real number line short of infinity—can be modeled using the approaches we describe in this paper.

## A Beta Way Is Possible

With the widespread availability of open-source software such as R (R Core Team, 2024) and its extensive ecosystem of user-developed packages, advanced models like beta regression have become increas-

<sup>1</sup>Specifically, for any continuous bounded variable  $x$ , we can rescale this variable to lie within 0 and 1 by using the formula  $x' = \frac{x - \min(x)}{\max(x) - \min(x)}$  where  $0 \leq x' \leq 1$ .

40 ingly accessible to applied researchers. Yet, their adoption in psychology remains relatively limited. One  
 41 contributing factor may be the lack of domain-specific examples that demonstrate how these models address  
 42 common challenges in psychological data. Although recent years have seen a growing interest in beta regres-  
 43 sion, and a number of useful tutorials are available (Bendixen & Purzycki, 2023; Coretta & Bürkner, 2025;  
 44 Heiss, 2021; e.g., Smithson & Verkuilen, 2006; Vuorre, 2019), these resources often either focus on basic  
 45 implementation or briefly mention extensions without detailing how they can be applied to psychologically  
 46 relevant research questions.

47 The present tutorial aims to help bridge this gap by providing a comprehensive, applied introduction  
 48 to beta regression and several of its extensions. In addition to the standard beta model, we walk through zero-  
 49 inflated, zero-one-inflated, and ordered beta regression. These models are particularly useful for researchers  
 50 working with proportion outcomes that include boundary values (e.g., exact 0s or 1s) or responses with an  
 51 inherent ordinal structure. Our goal is to offer practical guidance that enables psychological researchers to  
 52 implement, interpret, and report these models in ways that directly support their empirical questions.

53 Beyond model specification, we place strong emphasis on interpreting results on the response scale—  
 54 that is, in terms of probabilities and proportions—rather than relying on often difficult to interpret parameters.  
 55 This focus makes the models more accessible and meaningful for psychological applications, where effects  
 56 are often easier to communicate when framed on the original scale of the outcome (e.g., changes in recall  
 57 accuracy or task performance). Throughout, we provide reproducible code and annotated examples to help  
 58 readers implement and interpret these models in their own work.

59 We begin the tutorial with a non-technical overview of the beta distribution and its core parameters.  
 60 We then walk through the process of estimating beta regression models using the R package `brms` (Bürkner,  
 61 2017), illustrating each step with applied examples. To guide interpretation, we emphasize coefficients,  
 62 predicted probabilities, and marginal effects calculated using the `{marginaleffects}` package (Arel-Bundock  
 63 et al., 2024). We also introduce several useful extensions—zero-inflated (ZIB), zero-one-inflated (ZOIB), and  
 64 ordered beta regression—that enable researchers to model outcomes that include boundary values. Finally,  
 65 all code and materials used in this tutorial are fully reproducible and available via our GitHub repository:  
 66 [https://github.com/jgeller112/beta\\_regression\\_tutorial](https://github.com/jgeller112/beta_regression_tutorial)<sup>2</sup>.

## 67 Beta Distribution

68 Proportional data pose some challenges for standard modeling approaches: The data are bounded  
 69 between 0 and 1 and often exhibit non-constant variance (heteroscedasticity) (Ferrari & Cribari-Neto, 2004;  
 70 Paolino, 2001). Common distributions used within the generalized linear model frameworks often fail to  
 71 capture these properties adequately, which can necessitate alternative modeling strategies.

72 While we do not have time to delve fully into its derivation, the beta distribution is a preferred  
 73 distribution for this type of response because of certain unique properties. The beta distribution is defined as  
 74 a distribution of the uncertainty of probabilities, which must lie within 0 and 1. As a consequence, the beta  
 75 distribution is the maximum entropy distribution for any bounded continuous random variable, which means  
 76 that the beta distribution can represent the full range of possibilities of such a scale.<sup>3</sup> As a consequence, if we

---

<sup>2</sup>In this article, we try to limit code where possible; however, the online version has all the code needed to reproduce all analyses herein. Furthermore, to promote transparency and reproducibility, the tutorial was written in R version 4.5.1 (R Core Team (2024)) using Quarto (v.1.5.54), an open-source publishing system that allows for dynamic and static documents. This allows figures, tables, and text to be programmatically included directly in the manuscript, ensuring that all results are seamlessly integrated into the document. In addition, we use the `rix` (Rodrigues & Baumann, 2025) R package which harnesses the power of the `nix` (Dolstra & contributors, 2006) ecosystem to help with computational reproducibility. Not only does this give us a snapshot of the packages used to create the current manuscript, but it also takes a snapshot of system dependencies used at run-time. This way reproducers can easily re-use the exact same environment by installing the `nix` package manager and using the included `default.nix` file to set up the right environment. The README file in the GitHub repository contains detailed information on how to set this up to reproduce the contents of the current manuscript, including a video.

<sup>3</sup>Technically, this maximum entropy condition is satisfied because the  $\text{beta}(1,1)$  distribution is uniform over its support.

77 have a continuous scale with upper and lower bounds—and no other special conditions—the beta distribution  
 78 will in principle provide a very good approximation of the uncertainty of the scale.

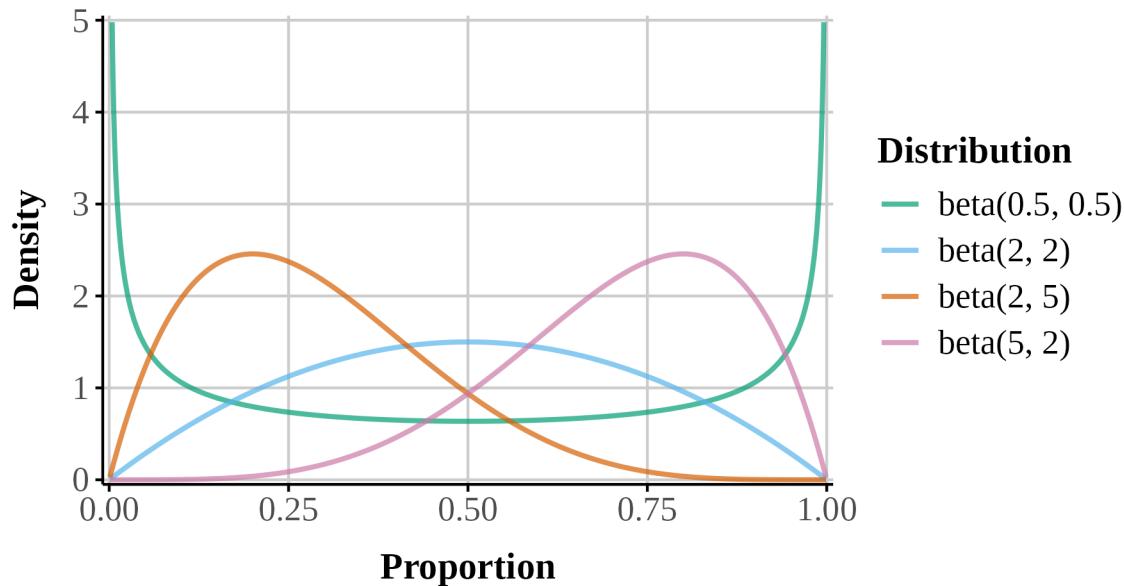
79 Typically, the expected value (or mean) of the response variable is the central estimand scholars want  
 80 to estimate. A model should specify how this expected value depends on explanatory variables through two  
 81 main components: a linear predictor, which combines the explanatory variables in a linear form ( $a + b_1x_1 +$   
 82  $b_2x_2$ , etc.), and a link function, which connects the expected value of the response variable to the linear  
 83 predictor (e.g.,  $E[Y] = g(a + b_1x_1 + b_2x_2)$ ). In addition, a random component specifies the distribution  
 84 of the response variable around its expected value (such as Poisson or binomial distributions, which belong  
 85 to the exponential family) (Nelder & Wedderburn, 1972). Together, these components provide a flexible  
 86 framework for modeling data with different distributional properties.

87 The beta distribution is continuous and restricted to values between 0 and 1 (exclusive). Its two  
 88 parameters—commonly called shape1 ( $\alpha$ ) and shape2 ( $\beta$ )—govern the distribution's location, skewness, and  
 89 spread. By adjusting these parameters, the distribution can take many functional forms (e.g., it can be sym-  
 90 metric, skewed, U-shaped, or even approximately uniform; see Figure 1).

91 To illustrate, consider a test question worth seven points. Suppose a participant scores five out of  
 92 seven. The number of points received (5) can be treated as  $\alpha$ , and the number of points missed (2) as  $\beta$ . The  
 93 resulting beta distribution would be skewed toward higher values, reflecting a high performance (yellow line  
 94 in Figure 1; “beta(5, 2)”). Reversing these values would produce a distribution skewed toward lower values,  
 95 representing poorer performance (green line in Figure 1; “beta(2, 5)”).

**Figure 1**

*beta distributions with different shape1 and shape2 parameters.*



## 96 I Can't Believe It's Not beta

97 While the standard parameterization of the beta distribution uses  $\alpha$  and  $\beta$ , a reparameterization to a  
 98 mean ( $\mu$ ) and precision ( $\phi$ ) is more useful for regression models. The mean represents the expected value  
 99 of the distribution, while the dispersion, which is inversely related to variance, reflects how concentrated  
 100 the distribution is around the mean, with higher values indicating a narrower distribution and lower values  
 101 indicating a wider one. The connections between the beta distribution's parameters are shown in Equation 1.

102 Importantly, the variance depends on the average value of the response because uncertainty intervals need to  
 103 adjust for how close the value of the response is to the boundary.

$$\begin{aligned} \text{Shape 1: } a &= \mu\phi & \text{Mean: } \mu &= \frac{a}{a+b} \\ \text{Shape 2: } b &= (1-\mu)\phi & \text{Precision: } \phi &= a+b \\ && \text{Variance: } var &= \frac{\mu \cdot (1-\mu)}{1+\phi} \end{aligned} \quad (1)$$

104 Thus, beta regression allows modeling both the mean and precision of the outcome distribution. To  
 105 ensure that  $\mu$  stays between 0 and 1, we apply a link function, which allows linear modeling of the mean on  
 106 an unbounded scale. A common link-function choice is the logit, but other functions such as the probit or  
 107 complementary log-log are possible.

108 The logit function,  $\text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$  links the mean to log-odds which are unbounded, making  
 109 linear modeling possible. The logit here no longer carries the same literal *odds* interpretation because there  
 110 are no corresponding counts of “successes” and “failures.” Instead, the logit transform here simply maps the  
 111 mean of the distribution to the real line. The inverse of the logit, called the logistic function, maps the linear  
 112 predictor  $\eta$  back to the original scale of the data  $\left(\mu = \frac{1}{1+e^{-\eta}}\right)$ . The coefficients describe how predictors shift  
 113 the *average proportion* on the logit scale here. Similarly, the strictly positive dispersion parameter is usually  
 114 modeled through a log link function, ensuring it remains positive.

115 By accounting for the observations’ natural limits and non-constant variance across different val-  
 116 ues, the beta distribution is useful in psychology where outcomes like performance rates or response scales  
 117 frequently exhibit these features.

## 118 Bayesian Approach to Beta Regression

119 Beta regression models can be estimated using either frequentist or Bayesian methods. In this paper,  
 120 we adopt a Bayesian framework because it facilitates the estimation and interpretation of more complex  
 121 models (Gelman et al., 2013; Johnson et al., 2022; McElreath, 2020). Additionally, the use of Bayesian statistics  
 122 in psychology has been steadily growing (**Pfadt2025?**). In principle, most Bayesian models can also be  
 123 estimated with frequentist approaches (e.g., maximum likelihood). However, as model complexity increases,  
 124 frequentist estimation often requires additional adjustments such as bootstrapping or custom optimization,  
 125 whereas Bayesian modeling handles these extensions more naturally through hierarchical structure and prior  
 126 specification.

127 A common concern is that Bayesian methods are slower than frequentist ones. While this is true  
 128 in general, modern Bayesian computation engines are efficient, and in explanatory modeling our priority  
 129 is to specify appropriate estimands rather than to optimize for computation speed. Moreover, we use the  
 130 `{brms}` package (Bürkner, 2017), which provides a high-level interface to Stan (Team, 2023) and uses familiar  
 131 R formula syntax, making advanced Bayesian modeling accessible to researchers with standard regression  
 132 backgrounds. The package also supports parallelization, which substantially reduces computation time for  
 133 large datasets.

134 There are several important differences between our Bayesian analysis and the frequentist methods  
 135 readers may be more familiar with—most notably, the absence of *t*- and *p*-values. To estimate models, the  
 136 `{brms}` package uses Stan’s computational algorithms to draw random samples from the posterior distribu-  
 137 tion, which represents uncertainty about the model parameters. This posterior is conceptually analogous to a  
 138 frequentist sampling distribution. By default, Bayesian models run 4 chains with 2,000 iterations each. The  
 139 first 1,000 iterations per chain are warmup and are discarded. The remaining 1,000 iterations per chain are  
 140 retained as posterior draws, yielding 4,000 total post-warmup draws across all chains. From these draws, we

141 can compute the posterior mean (analogous to a frequentist point estimate) and the 95% credible interval  
142 (Cr.I.), which is often compared to a confidence interval.

143 In addition, an important part of Bayesian analyses is prior specification. Priors encode our assumptions  
144 about plausible parameter values before observing the data and allow the model to regularize estimates,  
145 especially when data are sparse or parameters are weakly identified. To help bridge the conceptual gap for  
146 users more familiar with frequentist models, we begin with the default (weakly informative) priors provided  
147 by `{brms}`. These priors are intentionally modest, and in many applications produce results that closely  
148 align with frequentist estimates, while still offering the flexibility and interpretive advantages of a Bayesian  
149 framework. We strongly urge readers to consider prior specification strongly in all their work.

150 To ease readers into Bayesian data analysis we provide a metric known as the *probability of direction* (pd),  
151 which reflects the probability that a parameter is strictly positive or negative. When a uniform  
152 prior is used (all values equally likely in the prior), a pd of 95%, 97.5%, 99.5%, and 99.95% corresponds  
153 approximately to two-sided p-values of .10, .05, .01, and .001 (i.e.,  $\text{pd} \approx 1 - p/2$  for symmetric posteriors  
154 with weak/flat priors) (see Figure 2 for an illustrative comparison). For directional hypotheses, the pd can be  
155 interpreted as roughly equivalent to one minus the *p*-value (Marsman & Wagenmakers, 2016).

156 For reasons of space, we refer readers unfamiliar to Bayesian data analysis to several existing books  
157 on the topic (Gelman et al., 2013; Kruschke, 2015; McElreath, 2020). In addition, we assume readers are  
158 familiar with R, but those in need of a refresher should find Wickham et al. (2023) useful.

159

## Beta Regression Tutorial

### 160 Example Data

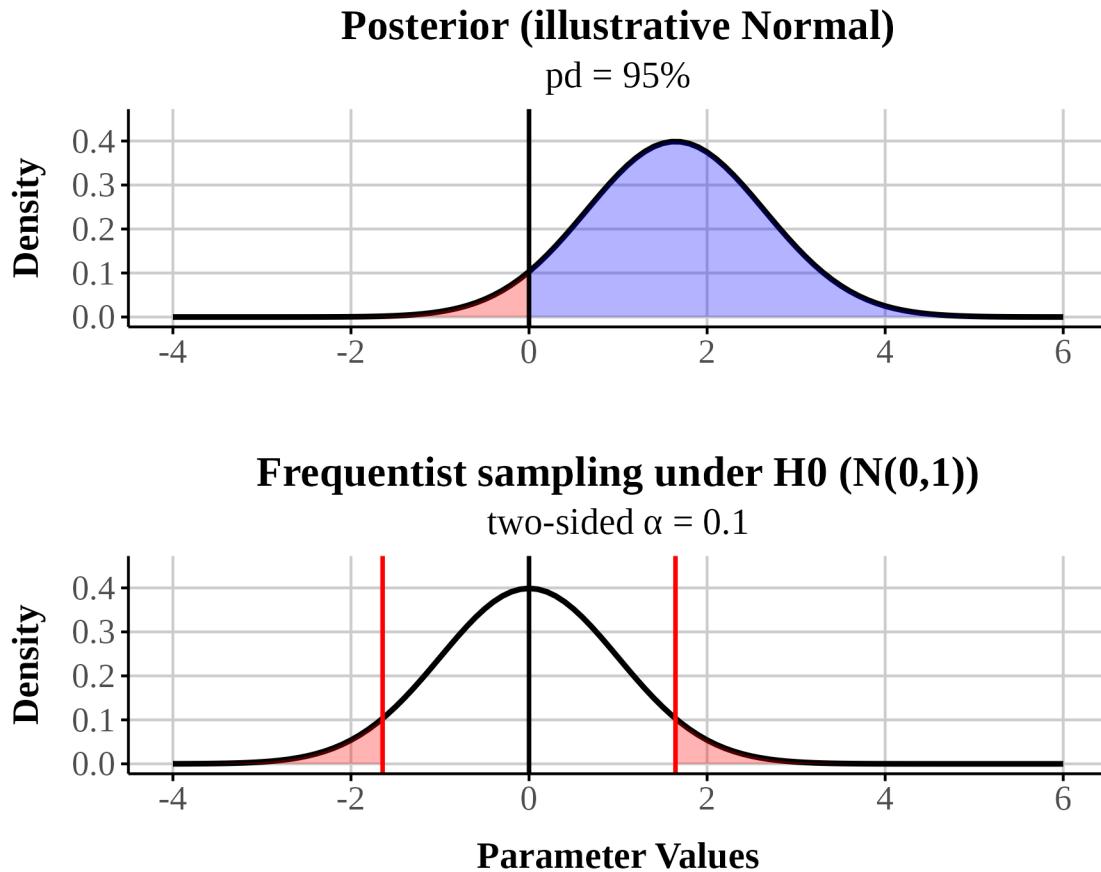
161 Throughout this tutorial, we analyze data from a memory experiment examining whether the flu-  
162 ency of an instructor's delivery affects recall performance (Wilford et al., 2020, Experiment 1A). Instructor  
163 fluency—marked by expressive gestures, dynamic vocal tone, and confident pacing—has been shown to  
164 influence students' perceptions of learning, often leading learners to rate fluent instructors more favorably  
165 (Carpenter et al., 2013). However, previous research suggests that these impressions do not reliably translate  
166 into improved memory performance (e.g., Carpenter et al., 2013; Toftness et al., 2017; Witherby & Car-  
167 penter, 2022). In contrast, Wilford et al. (2020) found that participants actually recalled more information  
168 after watching a fluent instructor compared to a disfluent one. This surprising finding makes the dataset a  
169 compelling case study for analyzing proportion data, as recall was scored out of 10 possible idea units per  
170 video.

171 In Experiment 1A, participants ( $N = 96$ ) watched two short instructional videos, each delivered  
172 either fluently or disfluently. Fluent videos featured instructors with smooth delivery and natural pacing,  
173 while disfluent videos included hesitations, monotone speech, and awkward pauses. After a distractor task,  
174 participants completed a free recall test, writing down as much content as they could remember from each  
175 video within a three-minute window. Their recall was then scored for the number of idea units correctly  
176 remembered.

177 Our primary outcome variable is the proportion of idea units recalled on the final test, calculated  
178 by dividing the number of correct units by 10. We show a sample of these data in Table 1. The dataset  
179 can be downloaded from GitHub (Listing 1). Because this is a bounded continuous variable (i.e., it ranges  
180 from 0 to 1), it violates the assumptions of typical linear regression models that treat outcomes as normally  
181 distributed. Despite this, it remains common in psychological research to analyze proportion data using  
182 models that assume normality. In what follows, we reproduce Wilford et al. (2020)'s analysis and then  
183 re-analyze the data using beta regression and highlight how it can improve our inferences.

**Figure 2**

A Bayesian posterior distribution (assuming a uniform prior) centered at a point estimate chosen so that the probability of direction ( $pd$ ) equals 0.95, and a frequentist sampling distribution (under the null; centered at 0). In the Bayesian posterior distribution, the blue area represents the  $pd$ , and the red area represents the remaining  $1 - pd$  of the distribution. In the frequentist sampling distribution, the red tail areas represent the rejection region at  $\alpha = 0.10$ . In this example, the posterior mean lies exactly at the  $1 - \frac{\alpha}{2}$  quantile of the null sampling distribution. For symmetric posteriors with flat priors, the  $pd$  is numerically equivalent to the one-sided  $p$ -value.

**Table 1**

Four observations from Wilford et al. (2020). Accuracy refers to the proportion of correctly recalled idea units.

Participant	Fluency	Accuracy
64	Disfluent	0.10
30	Fluent	0.60
12	Fluent	0.10
37	Fluent	0.35

---

**Listing 1** Data needed to run examples

---

```
# get data here from github
url <- str_glue(
  "https://raw.githubusercontent.com/jgeller112/",
  "beta_regression_tutorial/refs/heads/main/",
  "manuscript/data/fluency_data.csv"
)
fluency_data <- read.csv(url)
```

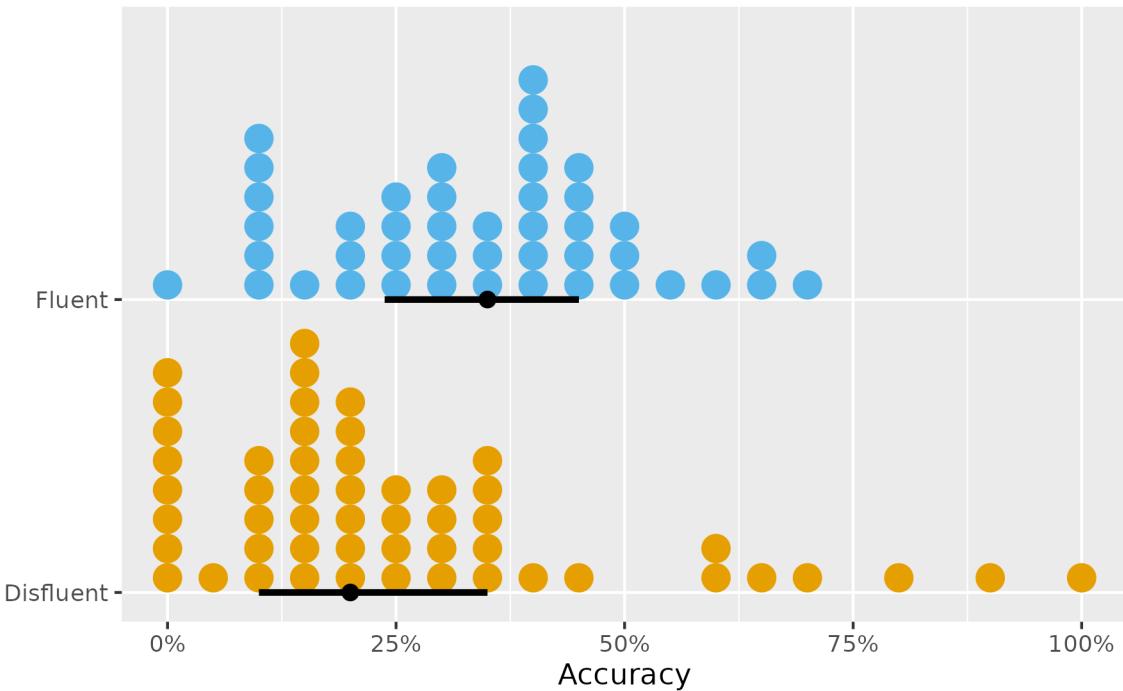
---

184 **Reanalysis of Wilford et al. Experiment 1A**

185 In their original analysis of Experiment 1A, Wilford et al. (2020) compared memory performance  
 186 between fluent and disfluent instructor conditions using a traditional independent-samples t-test on mean  
 187 accuracy for 96 participants. They found that participants who watched the fluent instructor recalled signifi-  
 188 cantly more idea units than those who viewed the disfluent version (see Figure 3).

**Figure 3**

*Dot plot depicting accuracy distributions for the Fluent and Disfluent conditions. Each condition shows individual data points and summary statistics to illustrate variability and central tendency.*



189 We first replicate this analysis in a regression framework using {brms}. We model final test mean  
 190 accuracy—the proportion of correctly recalled idea units across the videos—as the dependent variable. Our  
 191 predictor is instructor fluency, with two levels: Fluent and Disfluent. We use treatment (dummy) coding,  
 192 which is the default in R. This coding scheme sets the first level of a factor (in alphabetical order) as the  
 193 reference level. In this case, Disfluent is the reference, and the coefficient for Fluent reflects the contrast  
 194 between fluent and disfluent instructor conditions.

---

195 ***Regression Model***

196 We first start by loading the `brms` (Bürkner, 2017) and `cmdstanr` (Gabry et al., 2024) packages  
197 (Listing 2). We use the `cmdstanr` backend for Stan (Team, 2023) because it's faster than the default used to  
198 run models (i.e., `rstan`),<sup>4</sup> though all of these models can also be fit with `brms` defaults.

---

**Listing 2** Load the `brms` and `cmdstanr` packages

---

```
library(brms)
library(cmdstanr)
```

---



---

**Listing 3** Fitting a gaussian model with `brm()`.

---

```
bayes_reg_model <- brm(
  Accuracy ~ Fluency,
  data = fluency_data,
  family = gaussian(),
  file = "model_reg_bayes"
)
```

---

199 We fit the model using the `brm()` function from the `brms` package (Listing 3). Although not shown  
200 here, we ran the models using four chains (the default), executed in parallel across four cores. When the  
201 model is run in Listing 3, the model summary output will appear in the R console. The output from  
202 `bayes_reg_model` shows each parameter's posterior summary: The posterior distribution's mean and stan-  
203 dard deviation (analogous to the frequentist standard error) and its 95% credible interval, which indicate the  
204 95% of the most credible parameter values. In `brms`, the reported Cr.I is an equal-tailed interval, meaning  
205 that the probability mass excluded from the interval is split equally between the lower and upper tails. Ad-  
206 ditionally, the output indicates numerical estimates of the sampling algorithm's performance: Rhat should  
207 be close to one, and the ESS (effective sample size) metrics should be as large as possible given the number  
208 of iterations specified (default is 4000). Generally, ESS  $\geq 1000$  is recommended (Bürkner, 2017). For the  
209 models we present in this paper, convergence is trivial with standard linear models, though we note that these  
210 metrics are still important to pay attention to in case of model mis-fit.

211 Our main question of interest is: what is the causal effect of instructor fluency on final test perfor-  
212 mance? In order to answer this question, we will have to look at the output summary produced by Listing 3  
213 (also see Table 2 under Bayesian LM). the Intercept refers to the posterior mean accuracy in the disfluent  
214 condition,  $M = 0.256$ , as fluency was dummy-coded. The fluency coefficient (`FluencyFluent`) reflects the  
215 mean posterior difference in recall accuracy between the fluent and disfluent conditions:  $b = 0.084$ . The 95%  
216 Cr.I for this estimate spans from 0 to 0.169. These values are shown in the “95% Cr.I” columns of the output.  
217 These results closely mirror the findings reported by Wilford et al. (2020) (Experiment 1A).

```
218 Family: gaussian
219 Links: mu = identity; sigma = identity
220 Formula: Accuracy ~ Fluency
221 Data: fluency_data (Number of observations: 96)
```

---

<sup>4</sup>In order to use the `cmdstanr` backend you will need to first install the package (see <https://mc-stan.org/cmdstanr/>) and also run `cmdstanr::install_cmdstan()` if you have not done so already.

```

222
223 Regression Coefficients:
224             Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
225 Intercept      0.26      0.03    0.20    0.31 1.00     4034    3042
226 FluencyFluent  0.08      0.04    0.00    0.17 1.00     4367    2799
227
228 Further Distributional Parameters:
229             Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
230 sigma       0.21      0.02    0.18    0.24 1.00     3831    3161

```

The output also includes the effective sample size (ESS) and R (R-hat) values, both of which fall within acceptable ranges, indicating good model convergence. Throughout the tutorial, we focus primarily on posterior mean estimates and their 95% credible intervals. In addition, we report the pd measure in the main summary table (Table 2), provided by the `{bayestestR}` package (Makowski, Ben-Shachar, Chen, et al., 2019; Makowski, Ben-Shachar, & Lüdecke, 2019). This measure offers an intuitive parallel to *p*-values, which many readers may find familiar. For example, the fluency effect has a pd of .977, indicating a high probability that the effect is positive rather than negative (akin to  $p < .05$ ).

Importantly, pd does not indicate whether an effect is meaningfully different from a point value—it only reflects the proportion of the posterior in one direction. To address questions of practical significance, we encourage readers to consider the Region of Practical Equivalence (ROPE) with the Cr.Is (Kruschke, 2015). Unlike a hypothesis test of a point null (e.g., exactly zero), the ROPE defines a range of values that are deemed too small to be of substantive importance. As a rule of thumb (see Kruschke, 2018), if more than 95% of the posterior distribution lies inside the ROPE, the effect can be considered practically equivalent to that negligible range. If less than 5% lies inside, the effect can be considered meaningfully different. Intermediate cases are typically labeled undecided.

The `rope()` function in the `{bayestestR}` package computes the proportion of the posterior within the ROPE to facilitate this evaluation. By default, from bayesian models fit via `brms` the package determines a ROPE based on the data (roughly reflecting “negligible” effects), but these defaults should be used cautiously. The choice of ROPE should ultimately be guided by theoretical considerations, prior research, and the substantive context of the study. In `?@lst-rope`, we show how to compute this using `bayestestR`. Running the function with default settings suggests that less than 5% of the posterior distribution lies within the default ROPE (indicating the effect is larger than .02) (see Figure 4). Going forward we do not include a discussion of ROPE values, but we encourage readers to adopt it in their own research when appropriate.

Wilford et al. (2020) observed that instructor fluency impacts actual learning, using a standard *t*-test on the mean accuracy. But recall this approach assumes normality of residuals and homoscedacity. These assumptions are unrealistic when the response values approach the scale boundaries (Sladekova & Field, 2024). Does the data we have meet those assumptions? We can use the function `check_model()` from `easystats` (Lüdecke et al., 2022) to check our assumptions easily. The code in Listing 4 automatically produces Figure 5. We can see some issues with our data. Specifically, there appears to be violations of constant variance across the values of the scale (homoskedasticity). In plain terms, this type of model misspecification means that a standard OLS model can predict non-sensical values outside the bounds of the scale.

---

**Listing 4** Checking assumptions with the `check_model()` from `easystats` package .

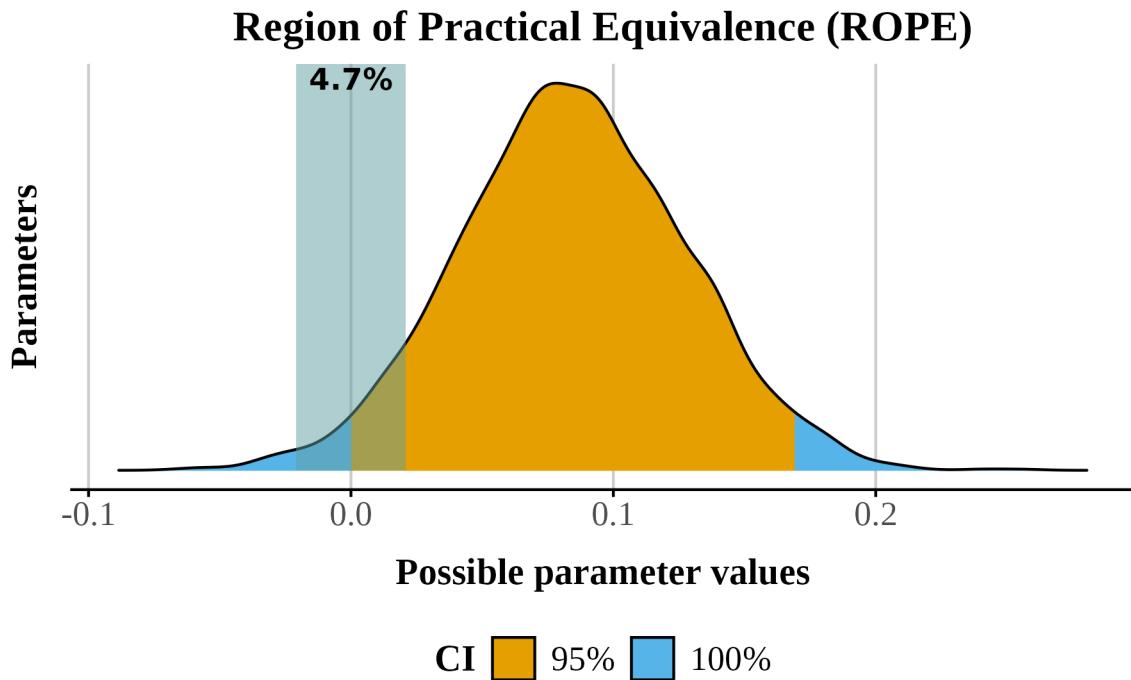
---

```
check_model(bayes_reg_model, check = c("homogeneity", "qq"))
```

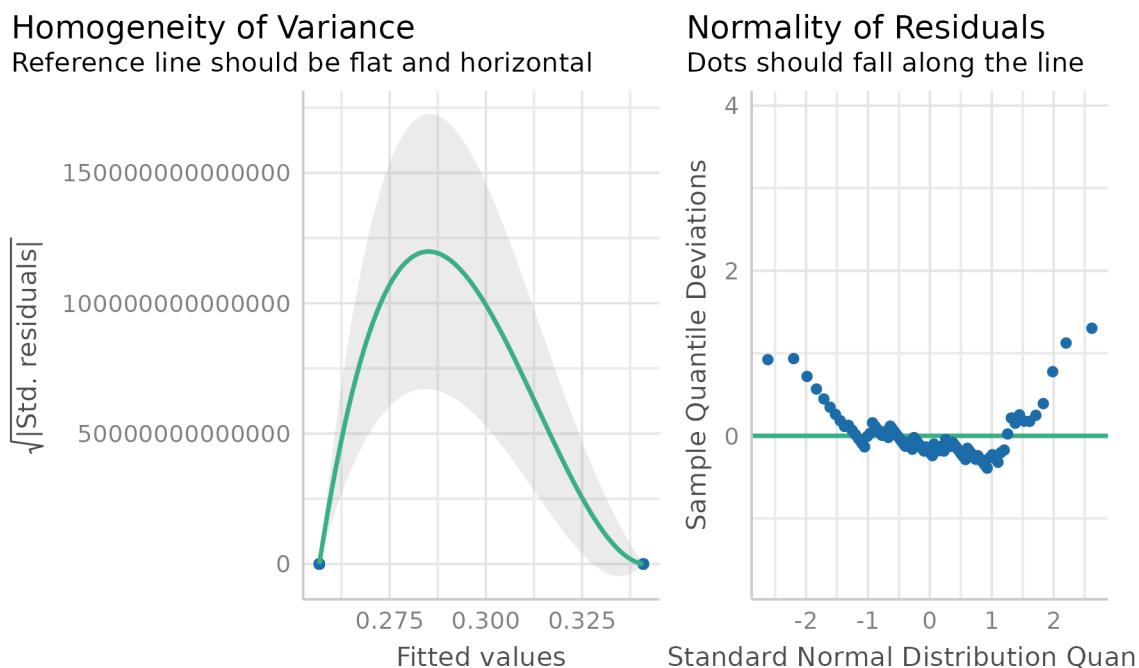
---

**Figure 4**

*Posterior distribution for the fluency effect showing the ROPE(shaded area) with 95% credible interval (orange) and 100% credible interval (blue). The percentage indicates the proportion of the posterior within the ROPE.*

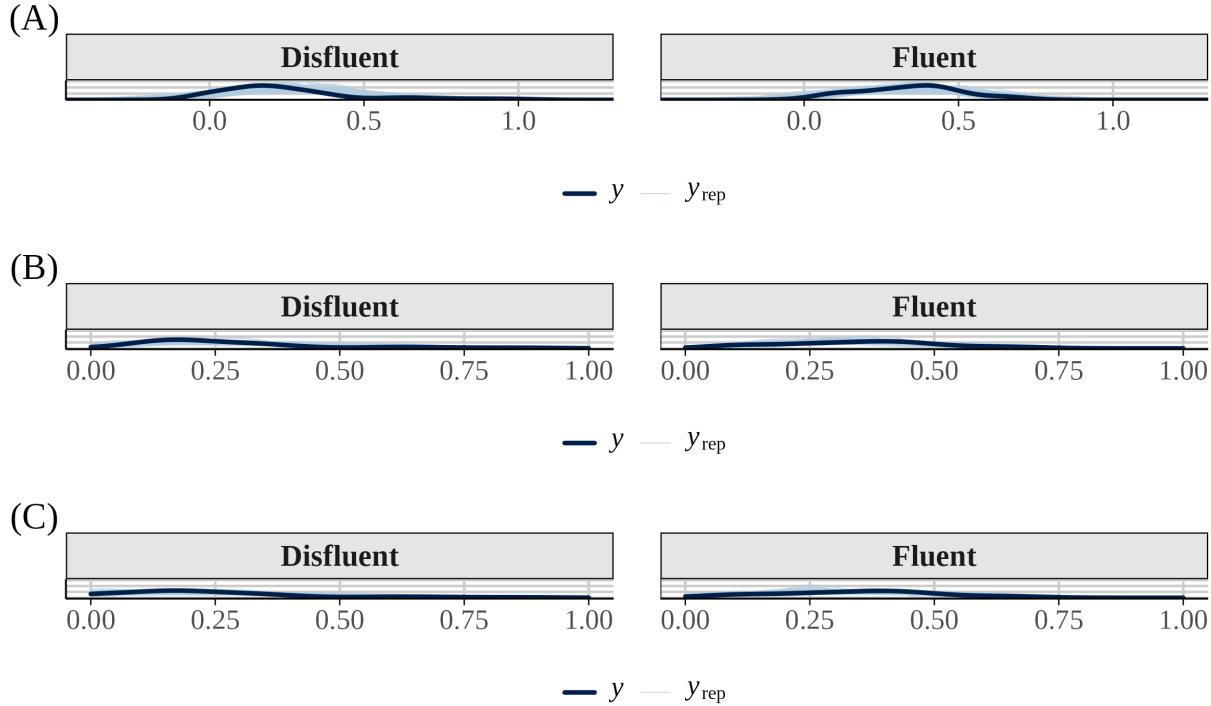
**Figure 5**

*Two assumption checks for our OLS model: Normality (left) and Homoskedasticity (right)*



**Figure 6**

The plots show 100 posterior predicted distributions with the label  $y_{\text{rep}}$  (light blue), the distribution of accuracy as function of fluency in dark blue for regular regression (A), beta regression (B), and ZIB (C) models



263 We can also examine how well the data fits the model by performing a posterior predictive check  
 264 using the `pp_check()` function from `brms`. A posterior predictive check involves looking at multiple draws  
 265 or repetitions from the posterior distribution and plotting it against the observed data. Ideally, the predictive  
 266 draws (the light blue lines) should show reasonable resemblance with the observed data (dark blue line). In  
 267 our example (see Figure 6 (A)) the model-predicted density is slightly too peaked and narrow compared to  
 268 the data. In addition, some of the draws extend into negative accuracy values.

## 269 Distributional Regression - Beta Regression

270 It is important to note that there are several justifiable approaches for addressing the distributional  
 271 issues observed in the data. For instance, one could analyze median accuracy instead of the mean, use some  
 272 type of robust estimator for heterogeneity, or apply non-parametric methods to relax some of the model  
 273 assumptions. However, in a Bayesian framework, we can address these issues more directly and transparently  
 274 by fitting distributional models ([kruschke2013?](#); [kneib2023?](#)). A key advantage of Bayesian distributional  
 275 modeling approach is that we are not limited to modeling only the mean or median of the outcome; we can  
 276 also model parameters such as the variance (or other shape parameters) as functions of predictors. This allows  
 277 us to examine how instructor fluency may influence not only average performance, but also the variability  
 278 in performance across students. If we wanted to keep our mean accuracy variable and continue to use a  
 279 Gaussian model, we could use a distributional approach and model the effect of fluency on  $\sigma$ .

280 Given the outcome variable is proportional, another solution would be to run a beta regression model.  
 281 Again, we can create the beta regression model in `brms`. In `brms`, we model each parameter independently.  
 282 Recall from the introduction that in a beta model we model two parameters— $\mu$  and  $\phi$ . Again we do this by

283 using the `bf()` function from `brms` (Listing 5). We fit two formulas, one for  $\mu$  and one for  $\phi$  and store it  
 284 in the `model_beta_bayes` object below. In the below `bf()` call, we are modeling Fluency as a function of  
 285 Accuracy only for the  $\mu$  parameter. For the  $\phi$  parameter, we are only modeling the intercept value. This is  
 286 saying dispersion does not change as a function of fluency.

287 To run our beta regression model, we need to exclude 0s and 1s in our data set. If we try to run  
 288 a model with our data `data_fluency` we get an error: `Error: Family 'beta' requires response`  
 289 `greater than 0`. This is because the beta distribution only supports observations in the 0 to 1 interval  
 290 *excluding exact 0s and 1s*. We need make sure there are no 0s and 1s in our dataset.

291 The dataset contains nine 0s and one 1. One approach is to nudge our 0s towards .01 and our 1s to  
 292 .99, or apply a special formula (Smithson & Verkuilen, 2006) so values fall within the [0, 1] interval. We  
 293 implore readers not to engage in this practice. Kubinec (2022) showed that this practice can result in serious  
 294 distortion of the outcome as the sample size grows larger, resulting in ever smaller values that are “nudged”.  
 295 Because the beta distribution is a non-linear model of the outcome, values that are very close to the boundary,  
 296 such as 0.00001 or 0.99999, will be highly influential outliers. To run this beta model we will remove the 0s  
 297 and 1s, and later in this article we will show how to jointly model these scale end points with the rest of the  
 298 data. The model from Listing 5 uses a transformed `data_fluency` object (called `data_beta`) where 0s and  
 299 1s are removed. When we run this code we should not get an error.

---

**Listing 5** Fitting a beta model without 0s and 1s in `brm()`.

---

```
# set up model formual
model_beta_bayes <- bf(
  Accuracy ~ Fluency, # fit mu model
  phi ~ 1 # fit phi model
)

# transform 0 to 0.1 and 1 to .99
data_beta <- fluency_data |>
  filter(
    Accuracy != 0,
    Accuracy != 1
  )

beta_brms <- brm(
  model_beta_bayes,
  data = data_beta,
  family = Beta(),
  file = "model_beta_bayes_reg_01"
)
```

---

300 **Model Parameters.** In Table 2, under the beta regression column, the coefficient with `b_` represents  
 301 how fluency of instructor influences the  $\mu$  parameter estimates (which is the mean of the distribution here).  
 302 These coefficients are interpreted on the scale of the logit, meaning they represent linear changes on a non-  
 303 linear space. The intercept term (`b_Intercept`) represents the log odds of the mean on accuracy for the  
 304 fluent instructor. Log odds that are negative indicate that it is more likely a “success” (like getting the correct  
 305 answer) will NOT happen than that it will happen. Similarly, regression coefficients in log odds forms that  
 306 are negative indicate that an increase in that predictor leads to a decrease in the predicted probability of a

**Table 2***Bayesian regression summaries for each model*

Parameter	Stat	Bayesian LM	Beta Regression	ZIB	ZOIB	Ordered Beta
b_Intercept	Mean	0.257	-0.830	-0.832	-0.831	-0.865
	Cr.I	[0.199, 0.315]	[-1.087, -0.55]	[-1.094, -0.552]	[-1.098, -0.559]	[-1.119, -0.596]
	pd	1.000***	1.000***	1.000***	1.000***	1.000***
b_Fluency	Mean	0.085	0.204	0.204	0.203	0.262
	Cr.I	[0.002, 0.166]	[-0.155, 0.539]	[-0.139, 0.545]	[-0.147, 0.541]	[-0.07, 0.598]
	pd	0.977*	0.875	0.872	0.880	0.936
sigma	Mean	0.209	-	-	-	-
	Cr.I	[0.179, 0.245]	-	-	-	-
	pd	1.000***	-	-	-	-
b_phi_Intercept	Mean	-	1.609	1.601	1.604	1.609
	Cr.I	-	[1.193, 2]	[1.187, 1.988]	[1.183, 1.989]	[1.179, 1.993]
	pd	-	1.000***	1.000***	1.000***	1.000***
b_phi_Fluency	Mean	-	0.420	0.425	0.426	0.408
	Cr.I	-	[-0.143, 0.993]	[-0.158, 0.994]	[-0.126, 0.994]	[-0.156, 0.983]
	pd	-	0.931	0.926	0.930	0.918
b_zi_Intercept	Mean	-	-	-1.673	-	-
	Cr.I	-	-	[-2.46, -0.978]	-	-
	pd	-	-	1.000***	-	-
b_zi_Fluency	Mean	-	-	-2.137	-	-
	Cr.I	-	-	[-4.618, -0.34]	-	-
	pd	-	-	0.992**	-	-
b_zoi_Intercept	Mean	-	-	-	-1.549	-
	Cr.I	-	-	-	[-2.339, -0.859]	-
	pd	-	-	-	1.000***	-
b_zoi_Fluency	Mean	-	-	-	-2.201	-
	Cr.I	-	-	-	[-4.449, -0.465]	-
	pd	-	-	-	0.996***	-
b_coi_Intercept	Mean	-	-	-	-2.022	-
	Cr.I	-	-	-	[-4.408, -0.287]	-
	pd	-	-	-	0.991**	-
b_coi_Fluency	Mean	-	-	-	0.245	-
	Cr.I	-	-	-	[-6.946, 5.716]	-
	pd	-	-	-	0.571	-

Note. Link functions: b\_mean = logit; b\_phi = log; b\_zoi (zero-one inflation) = logit; b\_coi (conditional one-inflation) = logit. Asterisks reflect approximate two-sided p-values derived from the posterior pd. pd  $\geq$  0.975 ( $p \leq .05$ ) = \*; pd  $\geq$  0.990 ( $p \leq .01$ ) = \*\*; pd  $\geq$  0.998 ( $p \leq .001$ ) = \*\*\*.

307 “success”.

308 The other component we need to pay attention to is the dispersion or precision parameter coefficients  
 309 labeled as phi in Table 2. The dispersion ( $\phi$ ) parameter tells us how precise our estimate is. Specifically,  
 310  $\phi$  in beta regression tells us about the variability of the response variable around its mean. Specifically, a  
 311 higher dispersion parameter indicates a narrower distribution, reflecting less variability. Conversely, a lower  
 312 dispersion parameter suggests a wider distribution, reflecting greater variability. The main difference between  
 313 a dispersion parameter and the variance is that the dispersion has a different interpretation depending on the  
 314 value of the outcome, as we show below. The best way to understand dispersion is to examine visual changes  
 315 in the distribution as the dispersion increases or decreases.

316 Understanding the dispersion parameter helps us gauge the precision of our predictions and the con-  
 317 sistency of the response variable. In `beta_brms` we only modeled the dispersion of the intercept. When  
 318  $\phi$  is not specified, the intercept is modeled by default (see Table 2). It represents the overall dispersion in  
 319 the outcome across all conditions. Instead, we can model different dispersions across levels of the Fluency  
 320 factor. To do so, we add Fluency to the phi model in `bf()`. We model the precision (phi) of the Fluency  
 321 factor by using a ~ and adding factors of interest to the right of it (Listing 6).

---

**Listing 6** Fitting beta model with dispersion in `brm()`.

---

```
model_beta_bayes_disp <- bf(
  Accuracy ~ Fluency, # Model of the mean
  phi ~ Fluency # Model of the precision
)

beta_brms_dis <- brm(
  model_beta_bayes_disp,
  data = data_beta,
  family = Beta(),
  file = "model_beta_bayes_dis_run01"
)
```

---

322 Table 2 displays the model summary with the precision parameter labeled as `phi_Fluency`. Since  $\phi$   
 323 is modeled on the log scale, the coefficients represent changes in  $\log-\phi$  rather than  $\phi$  itself. To see the effect  
 324 in the original units, we convert the values back by exponentiating. Thus, the effect of the Fluent condition  
 325 can be understood by comparing the exponentiated predicted  $\phi$  in the Fluent condition to that in the baseline  
 326 condition.

327 The  $\phi$  parameters are estimated on the log scale. The term  $\beta_{\text{Intercept}}^{(\phi)}$  represents the log-precision for  
 328 the reference (disfluent) condition. The coefficient  $\beta_{\text{FluencyFluent}}^{(\phi)}$  represents the change in log-precision when  
 329 moving from the disfluent to the fluent condition.

330 To obtain precision on the original scale, we exponentiate the linear predictor:

$$\phi_{\text{disfluent}} = \exp(\beta_{\text{Intercept}}^{(\phi)}), \quad \phi_{\text{fluent}} = \exp(\beta_{\text{Intercept}}^{(\phi)} + \beta_{\text{FluencyFluent}}^{(\phi)}).$$

331 The coefficient  $\beta_{\text{FluencyFluent}}^{(\phi)}$  therefore describes a *multiplicative* change in precision. Specifically,

$$\frac{\phi_{\text{fluent}}}{\phi_{\text{disfluent}}} = \exp(\beta_{\text{FluencyFluent}}^{(\phi)}).$$

<sup>332</sup> Because the 95% credible interval for  $\beta_{\text{FluencyFluent}}^{(\phi)}$  does not include zero, we infer that there is a  
<sup>333</sup> credible difference in precision between the fluent and disfluent conditions.

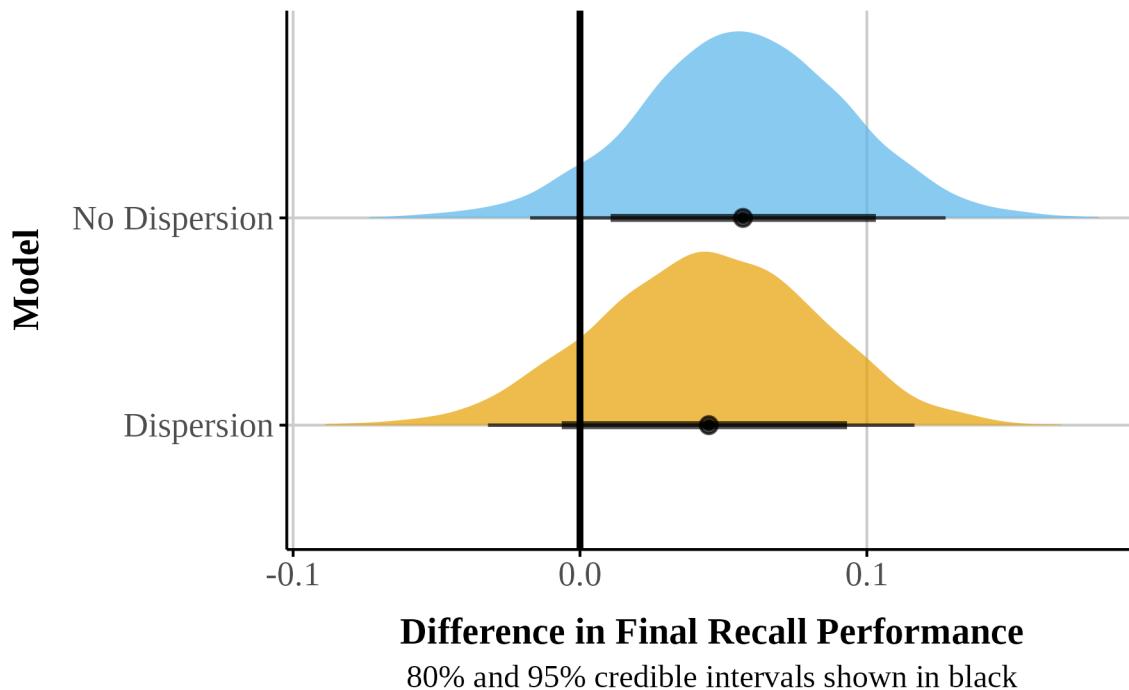
<sup>334</sup> It is important to note that these estimates are not the same as the marginal effects we discussed  
<sup>335</sup> earlier. Changes in dispersion affect the spread or variability of the response distribution without necessarily  
<sup>336</sup> altering its mean. This makes dispersion particularly relevant for research questions that focus on features  
<sup>337</sup> of the distribution beyond the average—such as how concentrated responses are. For instance, high dispersion  
<sup>338</sup> might indicate that individuals cluster at the extremes (e.g., very high or very low ratings), suggesting  
<sup>339</sup> clustering in the outcome.

<sup>340</sup> A critical assumption of the linear model is homoscedasticity, which means constant variance of the  
<sup>341</sup> errors. WIth beta regression model we can include a dispersion parameter for Fluency. Properly accounting  
<sup>342</sup> for dispersion is crucial because it impacts the precision of our mean estimates and, consequently, the sig-  
<sup>343</sup> nificance of our coefficients. The inclusion of dispersion in the our model increased the uncertainty of the  $\mu$   
<sup>344</sup> coefficient (see Figure 7). This suggests that failing to account for the dispersion of the variables might lead  
<sup>345</sup> to biased estimates. This highlights the potential utility of an approach like beta regression over a traditional  
<sup>346</sup> approach as beta regression can explicitly model dispersion and address issues of heteroscedasticity.

<sup>347</sup> While it is advisable to model precision, if there is uncertainty about the best model, a relatively  
<sup>348</sup> agnostic approach would be to compare models, for example with leave one out (loo) cross validation, to  
<sup>349</sup> examine if a dispersion parameter should be considered in our model.<sup>5</sup>

**Figure 7**

*Comparison of posterior distributions for the risk difference in fluency: Simple model (no dispersion for Fluency) vs. complex model with dispersion*



<sup>5</sup>The model fit statistic LOO-CV can be compared for any set of fitted brms models with the function loo().

**Table 3**

*Predicted probabilities for fluency factor.*

Fluency	Mean	95% Cr.I
Disfluent	0.297	[0.248, 0.348]
Fluent	0.354	[0.303, 0.407]

---

350 ***Predicted Probabilities***

351 Parameter estimates are usually difficult to interpret on their own and can require a lot of mathematical  
 352 gymnastics to get the estimate you need. We argue that researchers should not spend too much time  
 353 interpreting raw coefficients from non-linear models. We report them in this tutorial for completeness. In-  
 354 stead researchers should discuss the effects of the predictor on the actual outcome of interest (in this case the  
 355 0-1 scale). The logit link allows us to transform back and forth between the scale of a linear model and the  
 356 nonlinear scale of the outcome, which is bounded by 0 and 1. By using the inverse of the logit, we can eas-  
 357 ily transform our linear coefficients to obtain average effects on the scale of the proportions or percentages,  
 358 which is usually what is interesting to applied researchers. In a simple case, we can do this manually, but  
 359 when there are many factors in your model this can be quite complex.

360 In our example, we can use the `plogis()` function in base R to convert estimates from the logit scale  
 361 to the probability scale. The intercept of our model is -0.918, which reflects the logit of the mean accuracy  
 362 in the disfluent condition. If the estimated difference between the fluent and disfluent conditions is 0.24 on  
 363 the logit scale, we first add this value to the intercept value (-0.918) to get the logit for the fluent condition:  
 364  $-0.83 + 0.20 = -0.63$ . We then use `plogis()` to convert both logit values to probabilities (Fluent =  
 365 35%, Disfluent = 30%).

366 This is pretty easy to do manually, but when your model has many predictors, it can be quite cum-  
 367 bersome. To help us extract predictions from our model and visualize them we will use a package called  
 368 `{marginaleffects}` (Arel-Bundock et al., 2024) (see Listing 7). To get the proportions for each of our categor-  
 369 ical predictors on the  $\mu$  parameter we can use the function from the package called `predictions()`. These  
 370 are displayed in Table 3. These probabilities match what we calculated above.

---

**Listing 7** Load the `{marginaleffects}` package.

---

```
library(marginaleffects)
options(marginaleffects_posterior_center = mean) # make sure returns mean
```

---



---

**Listing 8** Predictions from the beta model for each level of Fluency.

---

```
predictions(
  beta_brms,
  # need to specify the levels of the categorical predictor
  newdata = datagrid(Fluency = c("Disfluent", "Fluent"))
)
```

---

371 For the Fluency factor, we can interpret Mean as proportions or percentages. That is, partic-  
 372 ipants who watched the fluent instructor scored on average 35% on the final exam compared to 30% for

373 those who watched the disfluent instructor. We can also visualize these from `{marginaleffects}` using the  
374 `plot_predictions()` function (see Listing 9).

---

**Listing 9** Plot predicted probabilities using `plot_predictions()` from `{marginaleffects}`


---

```
beta_plot <- plot_predictions(beta_brms, by = "Fluency")
```

---

375 The `plot_predictions()` function will only display the point estimate with the 95% credible inter-  
376 val. However, Bayesian estimation methods generate distributions for each parameter. This approach allows  
377 visualizing full uncertainty estimates beyond points and intervals. Using the `{marginaleffects}` package, we  
378 can obtain samples from the posterior distribution with the `posterior_draws()` function (see Listing 10).  
379 We can then plot these results to illustrate the range of plausible values for our estimates at different levels  
380 of uncertainty (see Figure 8).

---

**Listing 10** Extracting posterior draws from the beta regression model.

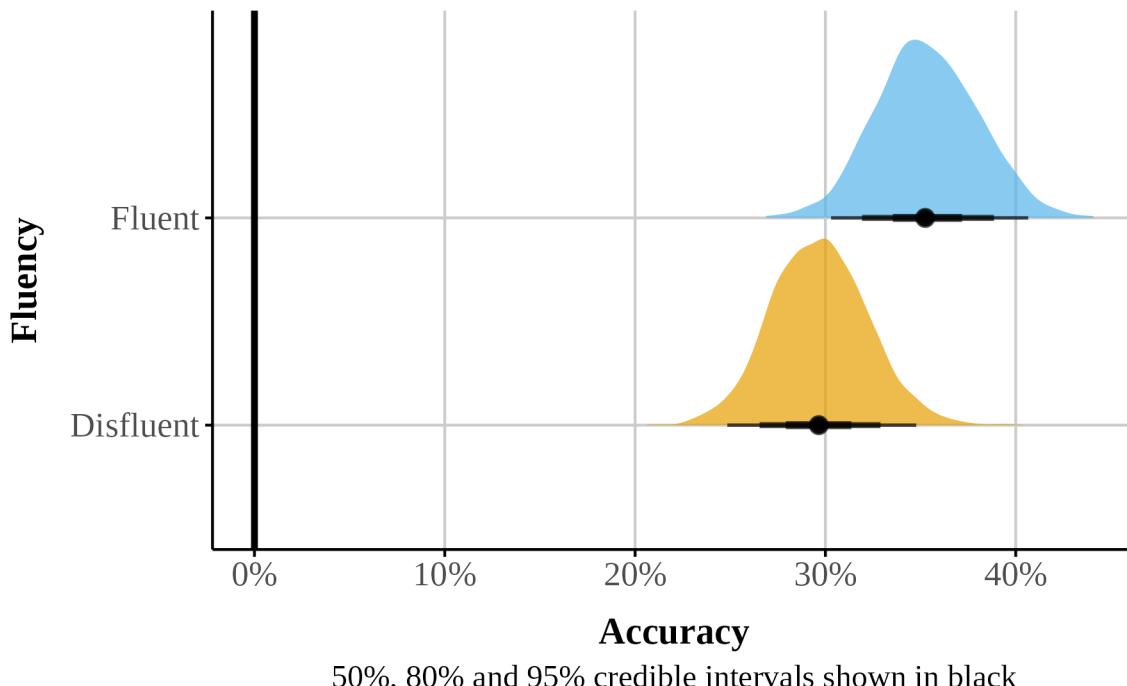
---

```
# Add a model identifier to each dataset
pred_draws_beta <- avg_predictions(beta_brms, variables = "Fluency") |>
  posterior_draws()
```

---

**Figure 8**

*Predicted probability posterior distributions by fluency*



**Table 4**

*Fluency difference*

Contrast	Mean	95% Cr.I
Fluent - Disfluent	0.057	[-0.017, 0.127]

### 381 **Marginal Effects**

382 Marginal effects offer an interpretable way to quantify how changes in a predictor influence an out-  
 383 come, while holding other factors constant in a specific manner. **In recent years, there has been a thrust to**  
 384 **move away from reporting regression coefficients alone, focusing instead on estimates that are easier to**  
 385 **interpret and communicate—particularly in non-linear models (McCabe et al., 2021; Rohrer & Arel-**  
 386 **Bundock, 2025).** Technically, marginal effects are computed as partial derivatives for continuous variables  
 387 or as finite differences for categorical (and sometimes continuous) predictors, depending on the structure  
 388 of the data and the research question. Substantively, these procedures translate raw regression coefficients  
 389 into quantities that reflect changes in the bounded outcome—for example, an  $x\%$  change in the value of a  
 390 proportion.

391 There are various types of marginal effects, and their calculation can vary across software packages.  
 392 For example, the popular `{emmeans}` package (Lenth, 2025) computes marginal effects by holding all pre-  
 393 dictors at their means (MEM). In this tutorial, we will use the `{marginaleffects}` package (Arel-Bundock et  
 394 al., 2024), which focuses on average marginal effects (AMEs) by default. AMEs summarize effects by gen-  
 395 erating predictions for each row of the original dataset and then averaging these predictions. This approach  
 396 retains a strong connection to the original data while offering a straightforward summary of the effect of  
 397 interest.

398 One practical use of AMEs is to estimate the average difference between two groups or conditions  
 399 which corresponds to the average treatment effect (ATE). Using the `avg_comparisons()` function in the  
 400 `{marginaleffects}` package (Listing 11), we can compute this quantity directly. By default, the function returns  
 401 the discrete difference between groups. When we take the difference in proportions between two groups it  
 402 is called the risk difference. Depending on the audience and modeling goals, the function can also produce  
 403 alternative effect size metrics, such as odds ratios or risk ratios. This flexibility makes it a powerful approach  
 404 for summarizing and communicating regression results.

---

#### **Listing 11** Calculating the difference between probabilities with `avg_comparisons()`

---

```
# get risk difference by default
beta_avg_comp <- avg_comparisons(beta_brms, comparison = "difference")
```

---

405 Table 4 presents the estimated difference for the Fluency factor (Mean column). The difference  
 406 between the fluent and disfluent conditions is 0.06, indicating that participants who watched a fluent instructor  
 407 scored, on average, 6% higher on the final recall test than those who watched a disfluent instructor. However,  
 408 the 95% credible interval includes 0 among the most credible values, suggesting we cannot rule out the  
 409 possibility of a null or weakly negative effect.

410 We can also use `{marginaleffects}` to get the actual precision difference between the two groups on  
 411  $\phi$  using similar code to above by setting `dpar` to “phi” {Listing 12}.

412 In psychology, it is common to report effect size measures like Cohen’s  $d$  (Cohen, 1977). When  
 413 working with proportions we can calculate something similar called Cohen’s  $h$ . Taking our proportions, we

---

**Listing 12** Calculating  $\phi$  difference with avg\_comparisons()

```
# get risk difference by default

beta_avg_phi <- avg_comparisons(
  beta_brms_dis,
  dpar = "phi",
  comparison = "difference"
)
```

---

414 can use the below equation (Equation 2) to calculate Cohen's  $h$  along with the 95% Cr.I around it. Using this  
 415 metric we see the effect size is small (0.107), 95% credible interval [-0.002, 0.361].

$$h = 2 \cdot (\arcsin(\sqrt{p_1}) - \arcsin(\sqrt{p_2})) \quad (2)$$

416 ***Posterior Predictive Check***

417 Figure 6 (B) shows the predictive check for our beta model. The model does a pretty good job at  
 418 capturing the data (The draws are now between 0-1) and the predicted values from the model follow the  
 419 observed data. However, it could be better.

420 **Zero-Inflated beta (ZIB) Regression**

421 A limitation of the beta regression model is that it can only accommodate values strictly between 0  
 422 and 1—it cannot handle values exactly equal to 0 or 1. In our dataset, we observed 9 rows where Accuracy  
 423 equals zero. To fit a beta regression model, we removed these values, but we have left out potentially valuable  
 424 information from our model—especially if the end points of the scale are distinctive in some way. In our case,  
 425 these 0s may be structural—that is, they represent real, systematic instances where participants failed to  
 426 answer correctly (rather than random noise or measurement error). For example, the fluency of the instructor  
 427 might be a key factor in predicting these zero responses. We will discuss two approaches for jointly modeling  
 428 these end points with the continuous data. First, we can use a zero-inflated beta (ZIB) model. This model  
 429 still estimates the mean ( $\mu$ ) and precision ( $\phi$ ) of the beta distribution for values between 0 and 1, but it also  
 430 includes an additional parameter,  $\alpha$ , which captures the probability of observing structural 0s.

431 The zero-inflated beta models a mixture of the data-generating process. The  $\alpha$  parameter uses a  
 432 logistic regression to model whether the data is 0 or not. Substantively, this could be a useful model when we  
 433 think that 0s come from a process that is relatively distinct from the data that is greater than 0. For example,  
 434 if we had a dataset with proportion of looks or eye fixations to certain areas on marketing materials, we might  
 435 want a separate model for those that do not look at certain areas on the screen because individuals who do  
 436 not look might be substantively different than those that look.

437 We can fit a ZIB model using brms() and use the {marginaleffects} package to make inferences  
 438 about our parameters of interest. Before we run a zero-inflated beta model, we will need to transform our  
 439 data again and remove the one 1 value in our data—we can keep our 0s. Similar to our beta regression model  
 440 we fit in brms, we will use the bf() function to fit several models. We fit our  $\mu$  and  $\phi$  parameters as well as  
 441 our zero-inflated parameter ( $\alpha$ ; here labeled as zi). In brms we can use the zero\_inflated\_beta family (see  
 442 Listing 13).

---

**Listing 13** Fitting zib model with `brm()`

---

```
# keep 0 but remove 1
data_beta_0 <- fluency_data |>
  filter(Accuracy != 1)

# set up model formula for zero-inflated beta in brm
zib_model <- bf(
  Accuracy ~ Fluency, # The mean of the 0-1 values, or mu
  phi ~ Fluency, # The precision of the 0-1 values, or phi
  zi ~ Fluency, # The zero-or-one-inflated part, or alpha
  family = zero_inflated_beta()
)

# fit zib model with brm
fit_zi <- brm(
  formula = zib_model,
  data = data_beta_0,
  file = "bayes_zib_model0not1"
)
```

---

443 **Posterior Predictive Check**

444 The ZIB model does a bit better at capturing the structure of the data than the beta regression model  
 445 (see Figure 6). Specifically, the ZIB model more accurately captures the increased density of values near  
 446 the lower end of the scale (i.e., near zero), which the standard beta model underestimates. The ZIB model's  
 447 predictive distributions also align more closely with the observed data across the entire range, particularly in  
 448 the peak and tail regions. This improved fit likely reflects the ZIB model's ability to explicitly model excess  
 449 0s (or near-zero values) via its inflation component, allowing it to better account for features in the data that  
 450 a standard beta distribution cannot accommodate.

451 **Predicted Probabilities and Marginal Effects**

452 Table 2, under the zero-inflated beta regression column, provides a summary of the posterior distribution  
 453 for each parameter. As stated before, it is preferable to back-transform our estimates to get probabilities.  
 454 To get the predicted probabilities we can again use the `avg_predictions()` and `avg_comparisons()`  
 455 functions from `{marginaleffects}` package (Arel-Bundock, 2024) to get predicted probabilities and the probability  
 456 difference between the levels of each factor. We can model the parameters separately using the `dpar`  
 457 argument setting to:  $\mu$ ,  $\phi$ ,  $\alpha$ . Here we look at the risk difference for Fluency under each parameter. If one  
 458 were interested in the average effect for the entire model, the `dpar` argument could be removed.

459 **Mu.** As shown in Table 5, there is little evidence for an effect of Fluency – the 95% Cr.I includes  
 460 zero, suggesting substantial uncertainty about the direction and magnitude of the effect—that is, though most  
 461 of the posterior density supports positive effects, nil and weakly negative effects cannot be ruled out.

462 **Dispersion.** As shown in Table 6, the posterior estimates suggest a credible effect of Fluency on  
 463 dispersion ( $\phi$ ), with disfluent responses showing greater variability. The 95% Cr.I for the fluency contrast  
 464 does not include zero, indicating a high probability in differences in precision.

**Table 5***Probability fluency difference ( $\mu$ )*

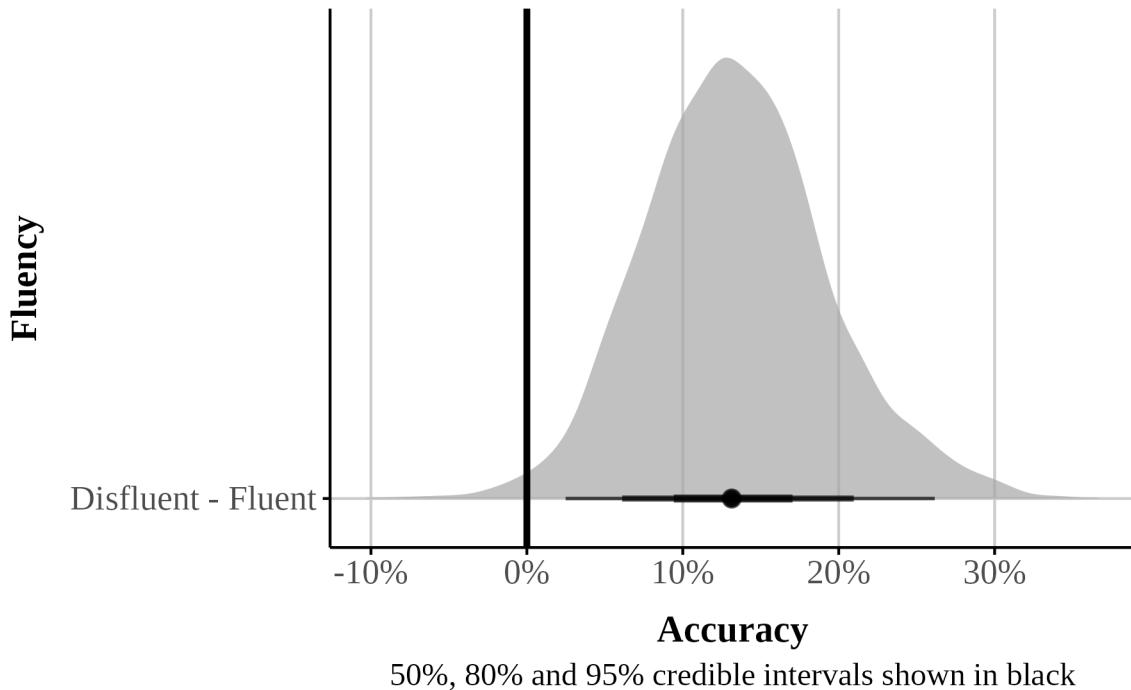
Contrast	Mean	95% Cr.I
Fluent - Disfluent	0.044	[-0.034, 0.118]

**Table 6***Probability fluency difference ( $\phi$ )*

Contrast	Mean	95% Cr.I
Fluent - Disfluent	2.7	[-0.847, 6.676]

465 **Zero-Inflation**

466 We can use {marginaleffects} to estimate and plot the posterior difference between the fluent and  
 467 disfluent conditions (see Figure 9). In Figure 9, the posterior distribution for this contrast lies mostly below  
 468 zero, indicating that a fluent instructor is associated with a lower probability of zero responses. The estimated  
 469 reduction is approximately 13%. The 95% credible interval does not include zero, which indicates that the  
 470 data provide consistent evidence for a reduction in zero responses under fluent instruction.

**Figure 9***Visualization of the predicted difference for zero-inflated part of model*

471 **Zero-One-Inflated beta (ZOIB)**

472 The ZIB model works well if you have 0s in your data, but not 1s.<sup>6</sup> In our previous examples we  
 473 either got rid of both 0s and 1s (beta regression), or removed the 1s (ZIB). Sometimes it is theoretically useful  
 474 to model both 0s and 1s as separate processes or to consider these values as essentially similar parts of the  
 475 continuous response, as we show later in the ordered beta regression model. For example, this is important  
 476 in visual analog scale data where there might be a prevalence of responses at the bounds (Kong & Edwards,  
 477 2016), in JOL tasks (Wilford et al., 2020), or in a free-list task where individuals provide open responses to  
 478 some question or topic which are then recoded to fall between 0-1 (Bendixen & Purzycki, 2023). Here 0s  
 479 and 1s are meaningful; 0 means item was not listed and 1 means the item was listed first.

480 Similar to the beta and zero-inflated models discussed above, we can fit a zero-and-one-inflated beta  
 481 (ZOIB) model in `{brms}` using the `zero_one_inflated_beta` family. This formulation simultaneously  
 482 estimates the mean  $\mu$  and precision  $\phi$  of the Beta component, as well as two inflation parameters:  $\alpha$ , the  
 483 probability that an observation is at either boundary (0 or 1), and  $\gamma$ , the conditional probability that, given  
 484 an observation falls on a boundary, it takes the value 1 rather than 0. In other words,  $\alpha$  determines how often  
 485 responses occur exactly at the endpoints, and  $\gamma$  determines the balance between zeros and ones among those  
 486 endpoint values. This specification allows the model to capture both the continuous variation in the interior  
 487 of the (0, 1) interval and the presence of exact boundary values.

488 To illustrate how  $\alpha$  and  $\gamma$  shape the distribution, Figure 10 displays simulated data across a range  
 489 of parameter combinations. As  $\alpha$  increases, more responses occur at the endpoints. As  $\gamma$  increases, the  
 490 proportion of those endpoint responses that are 1 increases relative to 0, producing increasingly pronounced  
 491 spikes at 1 as  $\gamma$  approaches 1. Together, these parameters give the ZOIB model the flexibility to represent  
 492 datasets with mixtures of continuous values and exact zeros and ones.

493 To fit a ZOIB model we use the `bf()` function. We model each parameter as a function of Fluency.  
 494 We then pass the `zoib_model` to our `brm()` function (see Listing 14). The summary of the output is in  
 495 Table 2 (under ZOIB).

496 **Model Parameters**

497 The output for the model is lengthy because we are estimating four distinct components, each with  
 498 their own independent responses and sub-models. All the coefficients are on the logit scale, except  $\phi$ , which is  
 499 on the log scale. Thankfully drawing inferences for all these different parameters, plotting their distributions,  
 500 and estimating their average marginal effects looks exactly the same—all the `brms` and `{marginaleffects}`  
 501 functions we used work the same.

502 **Predictions and Marginal Effects**

503 With `{marginaleffects}` we can choose `marginalize` over all the sub-models, averaged across the 0s,  
 504 continuous responses, and 1s in the data, or we can model the parameters separately using the `dpar` argument  
 505 like we did above setting it to:  $\mu$ ,  $\phi$ ,  $\alpha$ ,  $\gamma$  (see below). Using `avg_predictions()` and not setting `dpar` we  
 506 can get the predicted probabilities across all the sub-models. We can also plot the overall difference between  
 507 fluency and disfluency for the whole model with `plot_predictions()`.

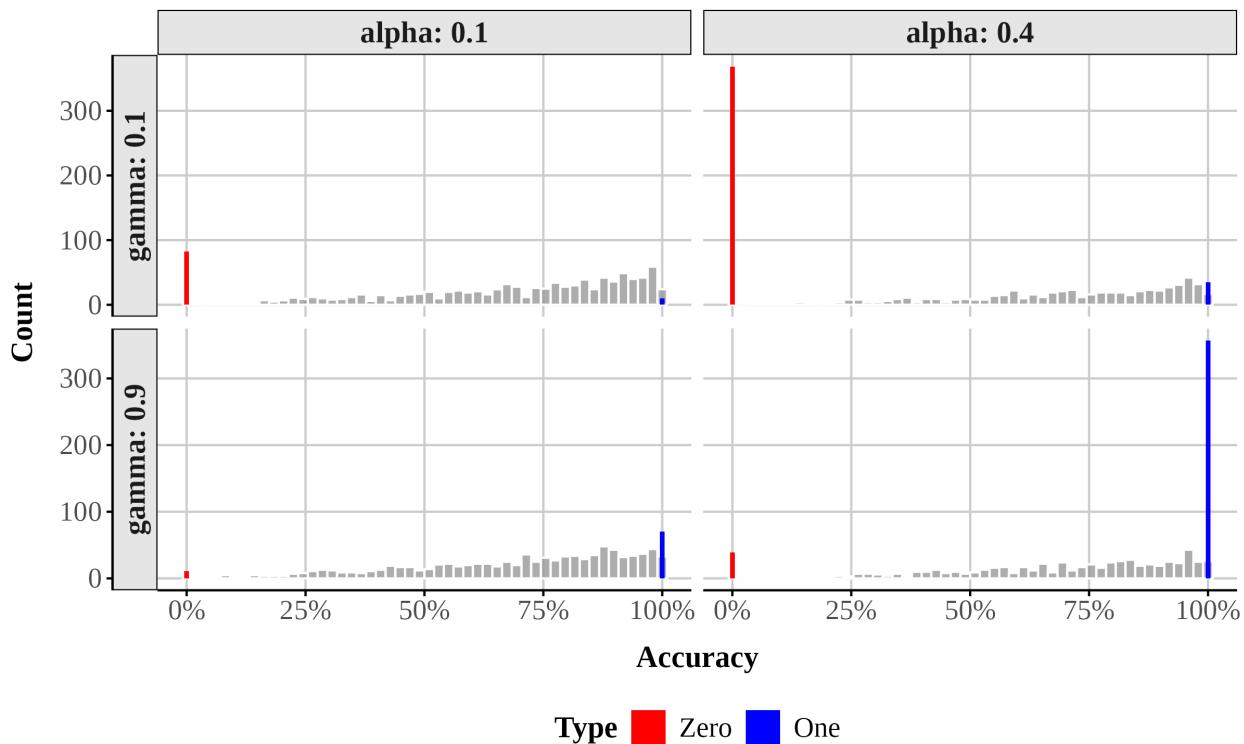
508 In addition, we show below how one can extract the predicted probabilities and marginal effects for  
 509  $\gamma$  (and a similar process for any other model component, `zoi`, etc.):

---

<sup>6</sup>In cases where your data include exact 1s but no 0s, you can fit a one-inflated beta regression model in `brms` by setting the `coi` parameter to 1. This tells the model that all point masses occur at 1, rather than being split between 0 and 1. In other words, `coi = 1` assumes that any inflation in the data is due entirely to values at 1. In our data, we have exactly one value equal to 1<sup>[^6]</sup>. While probably not significant to alter our findings, we can model 1s with a special type of model called the zero-one-inflated beta (ZOIB) model (Liu & Kong, 2015) if we believe that both 0s and 1s are distinct outcomes.

**Figure 10**

Simulated data from a ZOIB model illustrating the effects of the zero-one inflation parameter ( $\alpha$ ) and the conditional one-inflation parameter ( $\gamma$ ).



## 510 Ordered Beta Regression

511 Looking at the output from the ZOIB model (Table 2), we can see how running a model like this  
 512 can become fairly complex as it is fitting distinct sub-models for each component of the scale. The ability  
 513 to consider 0s and 1s as distinct processes from continuous values comes at a price in terms of complexity  
 514 and interpretability. A simplified version of the zero-one-inflated beta (ZOIB) model, known as ordered  
 515 beta regression (Kubinec, 2022; see also Makowski et al., 2025 for a reparameterized version called the  
 516 *beta-Gate* model), has been recently proposed. The ordered beta regression model exploits the fact that,  
 517 for most analyses, the continuous values (between 0-1) and the discrete outcomes (e.g., 0 or 1) are *ordered*.  
 518 For example, as a covariate  $x$  increases or decreases, we should expect the bounded outcome  $y$  to increase  
 519 or decrease monotonically as well from 0 to (0, 1) to 1. The ZOIB model does not impose this restriction;  
 520 a covariate could increase and the response  $y$  could increase in its continuous values while *simultaneously*  
 521 decreasing at *both* end points.<sup>7</sup> This complexity is not immediately obvious when fitting the ZOIB, nor is  
 522 it a potential relationship that many scholars want to consider when examining how covariates influence a  
 523 bounded scale.

524 To make the response ordered, the ordered beta regression model estimates a weighted combination  
 525 of a standard beta regression model for continuous responses and a logit model for the discrete values of  
 526 the response. By doing so, the amount of distinctiveness between the continuous responses and the discrete  
 527 end points is a function of the data (and any informative priors) rather than strictly defined as fully distinct  
 528 processes as in the ZOIB. For some datasets, the continuous and discrete responses will be fairly distinct,

<sup>7</sup>For a more complete description of this issue, we refer the reader to Kubinec (2022).

---

**Listing 14** Fitting a ZOIB model with `brm()`.

```
# fit the zoib model

zoib_model <- bf(
  Accuracy ~ Fluency, # The mean of the 0-1 values, or mu
  phi ~ Fluency, # The precision of the 0-1 values, or phi
  zoi ~ Fluency, # The zero-or-one-inflated part, or alpha
  coi ~ Fluency, # The one-inflated part, conditional on the 1s, or gamma
  family = zero_one_inflated_beta()
)

fit_zoib <- brm(
  formula = zoib_model,
  data = fluency_data,
  file = "bayes_zoib_model"
)
```

---

**Listing 15** Extracting predicted probabilities and marginal effects for conditional-one parameter

```
# get average predictions for coi param
coi_probs <- avg_predictions(fit_zoib, by = c("Fluency"), dpar = "coi")
# get difference between the two conditions
coi_me <- avg_comparisons(fit_zoib, variables = c("Fluency"), dpar = "coi")
```

---

529 and in others less so. To give a slightly absurd example, if a value of 0 meant that the subject was a fish,  
 530 and a value of 1 meant that the subject was now a trombone, then the ordered beta model would no longer  
 531 be appropriate.

532 The weights that average together the two parts of the outcome (i.e., discrete and continuous) are  
 533 determined by cutpoints that are estimated in conjunction with the data in a similar manner to what is known  
 534 as an ordered logit model. An in-depth explanation of ordinal regression is beyond the scope of this tutorial  
 535 (Bürkner & Vuorre, 2019; but see Fullerton & Anderson, 2021). At a basic level, ordinal regression models  
 536 are useful for outcome variables that are categorical in nature and have some inherent ordering (e.g., Likert  
 537 scale items). To preserve this ordering, ordinal models rely on the cumulative probability distribution.  
 538 Within an ordinal regression model it is assumed that there is a continuous but unobserved latent variable  
 539 that determines which of  $k$  ordinal responses will be selected. For example on a typical Likert scale from  
 540 ‘Strongly Disagree’ to ‘Strongly Agree’, you could assume that there is a continuous, unobserved variable  
 541 called ‘Agreement’.

542 While we cannot measure Agreement directly, the ordinal response gives us some indication about  
 543 where participants are on the continuous Agreement scale.  $k - 1$  cutoffs are then estimated to indicate the  
 544 point on the continuous Agreement scale at which your Agreement level is high enough to push you into the  
 545 next ordinal category (say Agree to Strongly Agree). Coefficients in the model estimate how much differ-  
 546 ent predictors change the estimated *continuous* scale (here, Agreement). Since there’s only one underlying  
 547 process, there’s only one set of coefficients to work with (proportional odds assumption). In an ordered beta  
 548 regression, three ordered categories are modeled: (1) exactly zero, (2) somewhere between zero and one,  
 549 and (3) exactly one. In an ordered beta regression, (1) and (2) are modeled with cumulative logits, where

550 one cutpoint is the the boundary between Exactly 0 and Between 0 and 1 and the other cutpoint is the bound-  
 551 ary between *Between 0 and 1* and *Exactly 1*. The continuous values in the middle, 0 to 1 (3), are modeled  
 552 as a vanilla beta regression with parameters reflecting the mean response on the logit scale as we have de-  
 553 scribed previously. Ultimately, employing cutpoints allows for a smooth transition between the bounds and  
 554 the continuous values, permitting both to be considered together rather than modeled separately as the ZOIB  
 555 requires.

556 The ordered beta regression model has shown to be more efficient and less biased than some of the  
 557 methods discussed (Kubinec, 2022) herein and has seen increasing use across the biomedical and social  
 558 sciences (Martin et al., 2024; Nouvian et al., 2023; Shrestha et al., 2024; Smith et al., 2024; Wilkes et al.,  
 559 2024) because it produces only a single set of coefficient estimates in a similar manner to a standard beta  
 560 regression or OLS.

### 561 **Fitting an Ordered Beta Regression**

562 To fit an ordered beta regression in a Bayesian context we use the `ordbetareg` (Kubinec, 2023) pack-  
 563 age. `ordbetareg` is a front-end to the `brms` package that we described earlier; in addition to the functions  
 564 available in the package, most `brms` functions and plots, including the diverse array of regression model-  
 565 ing options, will work with `ordbetareg` models. (We note that the `ordbeta` model is also available as a  
 566 maximum-likelihood variant in the R package `glmmTMB`.) We first load the `ordbetareg` package (see List-  
 567 ing 16).

---

#### **Listing 16 Load `ordbetareg`**

---

```
library(ordbetareg)
```

---

568 The `ordbetareg` package uses `brms` on the front-end so all the arguments we used previously apply  
 569 here. Instead of the `brm()` function we use `ordbetareg()`. To fit a model where dispersion does not vary  
 570 as a function of fluency we can use the below code (see Listing 17).

---

#### **Listing 17 Fitting ordered beta model with `ordbetareg()`**

---

```
ord_fit_brms <- ordbetareg(  

  Accuracy ~ Fluency,  

  data = fluency_data,  

  file = "bayes_ordbeta_model"  

)
```

---

571 However, if we want dispersion to vary as a function of fluency we can easily do that (see Listing 18).  
 572 Note the addition of the `phi_reg` argument in `m.phi`. This argument allows us to include a model that  
 573 explicitly models the dispersion parameter. Because we are modeling  $\phi$  as a function of fluency, we set the  
 574 the argument to both.

### 575 **Marginal Effects**

576 Table 2 presents the posterior summary. We can use `{marginaleffects}` to calculate differences on  
 577 the response scale that average over (or marginalize over) all our parameters.

**Listing 18** Fitting ordered beta model with dispersion using `ordbetareg()`


---

```
ord_beta_phi <- bf(Accuracy ~ Fluency, phi ~ Fluency)

m.phi <- ordbetareg(
  ord_beta_phi,
  data = fluency_data,
  phi_reg = 'both',
  file = "bayes_ordbeta_phi_model"
)
```

---

**Table 7**

*Marginal effect for fluency in ordered beta model*

Term	Contrast	Mean	95% Cr.I
Fluency	Fluent - Disfluent	0.062	[-0.013, 0.137]

578 In Table 7 the credible interval is close enough to zero relative to its uncertainty that we can conclude  
 579 there likely aren't differences between the conditions after taking dispersion and the 0s and 1s in our data  
 580 into account.

**581 Cutpoints**

582 The model cutpoints are not reported by default in the summary output, but we can access them with  
 583 the R package `posterior` (Bürkner et al., 2025) and the functions `as_draws` and `summary_draws`.

584 In Table 8, `cutzero` is the first cutpoint (the difference between 0 and continuous values) and `cutone`  
 585 is the second cutpoint (the difference between the continuous values and 1). These cutpoints are on the  
 586 logit scale and as such the numbers do not have a simple substantive meaning. In general, as the cutpoints  
 587 increase in absolute value (away from zero), then the discrete/boundary observations are more distinct from  
 588 the continuous values. This will happen if there is a clear gap or bunching in the outcome around the bounds.  
 589 This type of empirical feature of the distribution may be useful to scholars if they want to study differences  
 590 in how people perceive the ends of the scale versus the middle. It is possible, though beyond the scope of  
 591 this article, to model the location of the cutpoints with hierarchical (non-linear) covariates in `brms`. In the  
 592 most recent version of `ordbeta`, you can test the influence of different factors on these boundaries.

**593 Model Fit**

594 The best way to visualize model fit is to plot the full predictive distribution relative to the original  
 595 outcome. Because ordered beta regression is a mixed discrete/continuous model, a separate plotting function,

**Table 8**

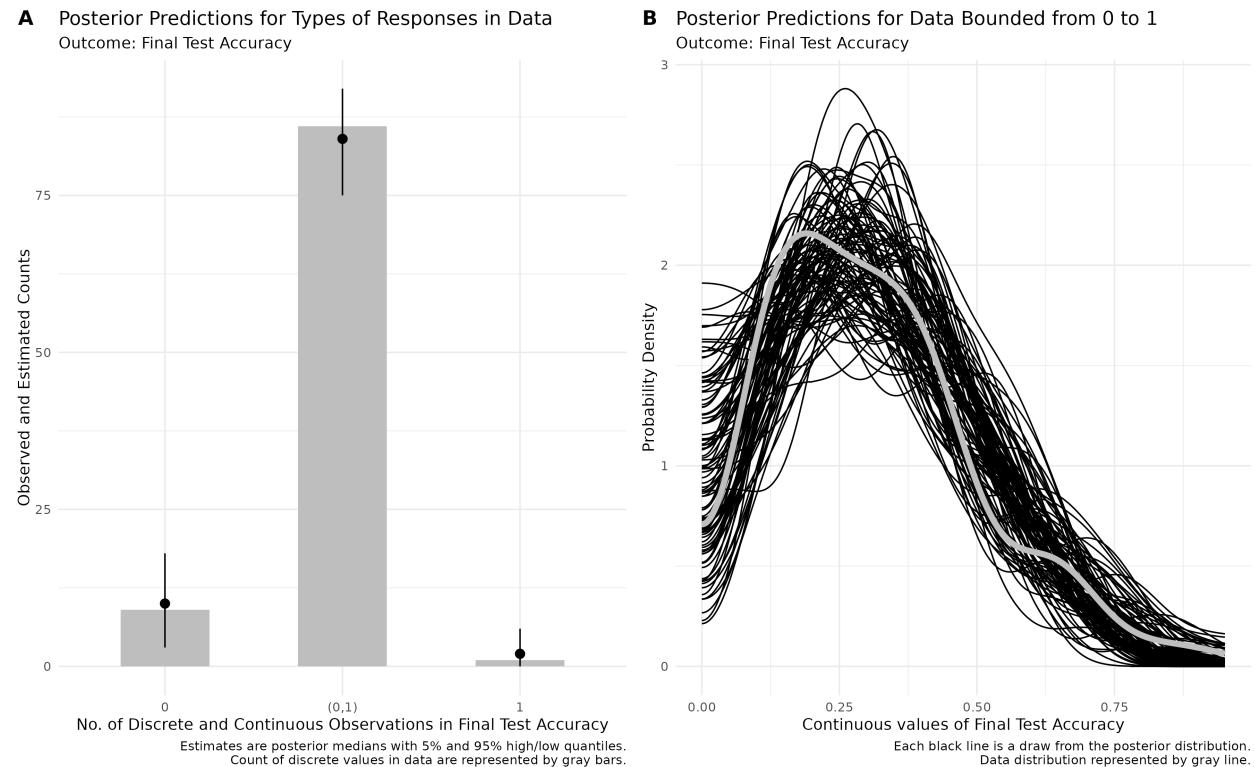
*Cutzero and cutone parameter summary*

Parameter	Mean	95% Cr.I
cutzero	-2.97	[-3.56, -2.43]
cutone	1.85	[1.64, 2.07]

596 `pp_check_ordbetareg`, is included in the `ordbetareg` package that accurately handles the unique features  
 597 of this distribution. The default plot in `brms` will collapse these two features of the outcome together, which  
 598 will make the fit look worse than it actually is. The `ordbetareg` function returns a list with two plots,  
 599 `discrete` and `continuous`, which can either be printed and plotted or further modified as `ggplot2` objects  
 600 (see Figure 11).

**Figure 11**

*Posterior predictive check for ordered beta regression model. A. Discrete posterior check. B. Continuous posterior check.*



601 The discrete plot, which is a bar graph, shows that the posterior distribution accurately captures the  
 602 number of different types of responses (discrete or continuous) in the data. For the continuous plot shown as  
 603 a density plot with one line per posterior draw, the model does a very good job at capturing the distribution.

604 Overall, it is clear from the posterior distribution plot that the ordered beta model fits the data well.  
 605 To fully understand model fit, both of these plots need to be inspected as they are conceptually distinct.

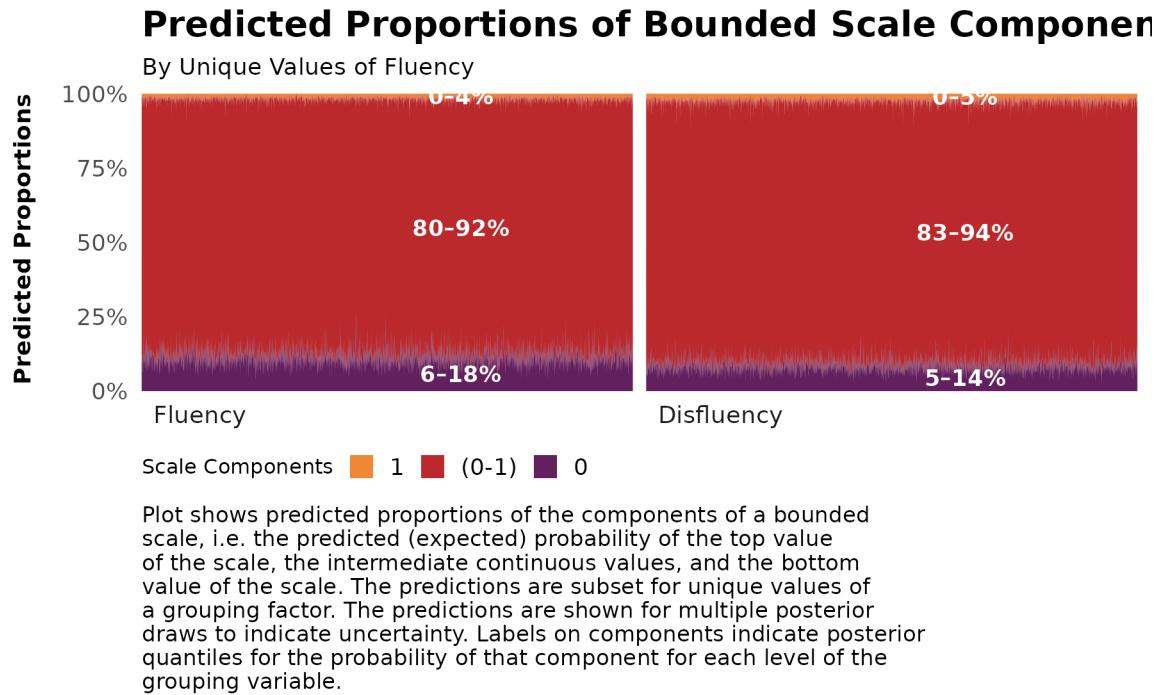
606 **Model Visualization.** `ordbetareg` provides a neat visualization function called `plot_heiss()`  
 607 (Ye & Heiss, 2023) that can represent dispersion in the entire outcome as a function of discrete covariates.  
 608 This function produces a plot of predicted proportions across the range of our Fluency factor. In Figure 12  
 609 we get predicted proportions for Fluency across the bounded scale. Looking at the figure we can see there is  
 610 much overlap between instructors in the middle portion ( $\mu$ ). However, we do see some small differences at  
 611 the zero bounds.

### 612 *Ordered Beta Scale*

613 In the `ordbetareg` function there is a `true_bound` argument. In cases where your data is not  
 614 bounded between 0-1, you can use this argument to specify the bounds of the argument to fit the ordered beta

**Figure 12**

*Heiss plot of predicted probabilities across the scale (0-100)*



615 regression. For example, your data might be bounded between 1 and 7. If so, you can model it as such and  
 616 `ordbetareg` will convert the model predictions back to the true bounds after estimation.

617

## Discussion

618 The use of beta regression in psychology, and the social sciences in general, is rare. With this tutorial,  
 619 we hope to turn the tides. Beta regression models are an attractive alternative to models that impose unrealistic  
 620 assumptions like normality, linearity, homoscedasticity, and unbounded data. Beyond these models, there  
 621 are a diverse array of different models that can be used depending on your outcome of interest.

622 Throughout this tutorial our main aim was to help guide researchers in running analyses with pro-  
 623 portional or percentage outcomes using beta regression and some of its alternatives. In the current example,  
 624 we used real data from Wilford et al. (2020) and discussed how to fit these models in R, interpret model  
 625 parameters, extract predicted probabilities and marginal effects, and visualize the results.

626 Comparing our analysis with that of Wilford et al. (2020), we demonstrated that using traditional  
 627 approaches (e.g., t-tests) to analyze accuracy data can lead to inaccurate inferences. Although we successfully  
 628 reproduced one of their key findings, our use of beta regression and its extensions revealed important nuances  
 629 in the results. With a traditional beta regression model—which accounts for both the mean and the precision  
 630 (dispersion)—we observed similar effects of instructor fluency on performance. However, the standard beta  
 631 model does not accommodate boundary values (i.e., 0s and 1s).

632 When we applied a ZIB model, which explicitly accounts for structural 0s, we found no effect of  
 633 fluency on the mean ( $\mu$ ) part of the model. Instead, the effect of fluency emerged in the structural zero  
 634 (*inflated zero;  $\alpha$* ) component. This pattern was consistent when using a zero-one-inflated beta (ZOIB) model.  
 635 Furthermore, we fit an ordered beta regression model (Kubinec, 2022), which appropriately models the full  
 636 range of values, including 0s and 1s. Here, we did not observe a reliable effect of fluency on the mean once

637 we accounted for dispersion.

638 These analyses emphasize the importance of fitting a model that aligns with the nature of the data.  
639 The simplest and recommended approach when dealing with data that contains 0s and/or 1s is to fit an ordered  
640 beta model, assuming the process is truly continuous. However, if you believe the process is distinct in nature,  
641 a ZIB or ZOIB model might be a better choice. Ultimately, this decision should be guided by theory.

642 For instance, if we believe fluency influences the structural zero part of the model, we might want to  
643 model this process separately using a ZIB or ZOIB. With the current dataset, fluency might affect specific  
644 aspects of performance (such as the likelihood of complete failure) rather than general performance levels.  
645 This effect could be due to participant disengagement during the disfluent lecture. If students fail to pay  
646 attention because of features of disfluency, they may miss relevant information, leading to a floor effect at  
647 the test. If this is the case, we would want to model this appropriately. However, if we believe fluency effects  
648 general performance levels, a model that takes in to account the entire process accounting for the 0s and 1s  
649 might be appropriate.

650 In the discussion section of Wilford et al. (2020), they were unable to offer a tenable explanation for  
651 performance differences based on instructor fluency. A model that accounts for the excess 0s in the dataset  
652 provides one testable explanation: watching a disfluent lecture may lead to lapses in attention, resulting  
653 in poorer performance in that group. These lapses, in turn, contribute to the observed differences in the  
654 fluent condition. This modeling approach opens a promising avenue for future research—one that would have  
655 remained inaccessible otherwise.

656 Not everyone will be eager to implement the techniques discussed herein. In such cases, the key  
657 question becomes: What is the least problematic approach to handling proportional data? One reasonable  
658 option is to fit multiple models tailored to the specific characteristics of your data. For example, if your data  
659 contain 0s, you might fit two models: a traditional linear model excluding the 0s, and a logistic model to  
660 account for the zero versus non-zero distinction. If your data contain both 0s and 1s, you could fit separate  
661 models for the 0s and 1s in addition to the OLS model. There are many defensible strategies to choose from  
662 depending on the context. However, we do not recommend transforming the values of your data (e.g., 0s to  
663 .01 and 1s to .99) or ignoring the properties of your data simply to fit traditional statistical models.

664 In this tutorial, we demonstrated how to analyze these models from a Bayesian perspective. While we  
665 recognize that not everyone identifies as a Bayesian, implementing these models using a Bayesian framework  
666 is relatively straightforward—it requires only a single package, lowering the barrier to entry. For those who  
667 prefer frequentist analyses, several R packages are available. For example, the `{betareg}` package (Cribari-  
668 Neto & Zeileis, 2010) `{glmmTMB}` (Brooks et al., 2017) and `{gamlss}` (2005) are nice options. To this end,  
669 I have included supplemental materials demonstrating how to use frequentist packages to analyze the data  
670 presented herein.

## 671 Conclusion

672 Overall, this tutorial emphasizes the importance of modeling the data you have. Although the ex-  
673 ample provided is relatively simple (a one-factor model with two levels), we hope it demonstrates that even  
674 with a basic dataset, there is much nuance in interpretation and inference. Properly modeling your data  
675 can lead to deeper insights, far beyond what traditional measures might offer. With the tools introduced in  
676 this tutorial, researchers now have the means to analyze their data effectively, uncover patterns, make ac-  
677 curate predictions, and support their findings with robust statistical evidence. By applying these modeling  
678 techniques, researchers can improve the validity and reliability of their studies, ultimately leading to more  
679 informed decisions and advancements in their respective fields.

680

## References

- 681 Arel-Bundock, V. (2024). *Marginaleffects: Predictions, comparisons, slopes, marginal means, and hypothesis tests*. <https://CRAN.R-project.org/package=marginaleffects>
- 682 Arel-Bundock, V., Greifer, N., & Heiss, A. (2024). How to interpret statistical models using marginaleffects for R and Python. *Journal of Statistical Software*, 111(9), 1–32. <https://doi.org/10.18637/jss.v111.i09>
- 683 Bartlett, M. S. (1936). The Square Root Transformation in Analysis of Variance. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 3(1), 68–78. <https://doi.org/10.2307/2983678>
- 684 Bendixen, T., & Purzycki, B. G. (2023). Cognitive and cultural models in psychological science: A tutorial on modeling free-list data as a dependent variable in Bayesian regression. *Psychological Methods*. <https://doi.org/10.1037/met0000553>
- 685 Brooks, M. E., Kristensen, K., van, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). *{glmmTMB} balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling*. 9. <https://doi.org/10.32614/RJ-2017-066>
- 686 Bürkner, P.-C. (2017). *{Brms}: An {r} package for {bayesian} multilevel models using {stan}*. 80. <https://doi.org/10.18637/jss.v080.i01>
- 687 Bürkner, P.-C., Gabry, J., Kay, M., & Vehtari, A. (2025). *posterior: Tools for working with posterior distributions*. <https://mc-stan.org/posterior/>
- 688 Bürkner, P.-C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, 2(1), 77–101. <https://doi.org/10.1177/2515245918823199>
- 689 Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*, 20(6), 1350–1356. <https://doi.org/10.3758/s13423-013-0442-z>
- 690 Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*, rev. ed. Lawrence Erlbaum Associates, Inc.
- 691 Coretta, S., & Bürkner, P.-C. (2025). *Bayesian beta regressions with brms in r: A tutorial for phoneticians*. [https://doi.org/10.31219/osf.io/f9rqg\\_v1](https://doi.org/10.31219/osf.io/f9rqg_v1).
- 692 Costello, T. H. et al. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385, eadq1814. <https://doi.org/10.1126/science.adq1814>
- 693 Cribari-Neto, F., & Zeileis, A. (2010). *Beta regression in {r}*. 34. <https://doi.org/10.18637/jss.v034.i02>
- 694 Dolstra, E., & contributors, T. N. (2006). *Nix* [Computer software]. <https://nixos.org/>
- 695 Ferrari, S., & Cribari-Neto, F. (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, 31(7), 799–815. <https://doi.org/10.1080/0266476042000214501>
- 696 Fullerton, A. S., & Anderson, K. F. (2021). Ordered Regression Models: a Tutorial. *Prevention Science*, 24(3), 431–443. <https://doi.org/10.1007/s11121-021-01302-y>
- 697 Gabry, J., Češnovar, R., Johnson, A., & Brodner, S. (2024). *Cmdstanr: R interface to 'CmdStan'*. <https://mc-stan.org/cmdstanr/>
- 698 Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (Third). CRC. <https://stat.columbia.edu/~gelman/book/>
- 699 Heiss, A. (2021). *A guide to modeling proportions with bayesian beta and zero-inflated beta regression models*. <http://dx.doi.org/10.59350/7p1a4-0tw75>
- 700 James, A. N., Ryskin, R., Hartshorne, J. K., Backs, H., Bala, N., Barcenas-Meade, L., Bhattacharai, S., Charles, T., Copoulos, G., Coss, C., Eisert, A., Furuhashi, E., Ginell, K., Guttman-McCabe, A., Harrison, E. (Chaz)., Hoban, L., Hwang, W. A., Iannetta, C., Koenig, K. M., ... Leeuw, J. R. de. (2025). What Paradigms Can Webcam Eye-Tracking Be Used For? Attempted Replications of Five Cognitive Science Experiments. *Collabra: Psychology*, 11(1). <https://doi.org/10.1525/collabra.140755>
- 701 Johnson, A., Ott, M., & Dogucu, M. (2022). *Bayes rules!: An introduction to applied bayesian modeling*.

- Routledge & CRC Press.
- Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, 59, 40–57. <https://doi.org/10.1016/j.wocn.2016.08.006>
- Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories. *Psychological Science*, 19(6), 585–592. <https://doi.org/10.1111/j.1467-9280.2008.02127.x>
- Kruschke, J. K. (2015). *Doing bayesian data analysis: A tutorial with r, JAGS, and stan* (2nd ed.). Academic Press.
- Kruschke, J. K. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. *Advances in Methods and Practices in Psychological Science*, 1(2), 270–280. <https://doi.org/10.1177/2515245918771304>
- Kubinec, R. (2022). Ordered Beta Regression: A Parsimonious, Well-Fitting Model for Continuous Data with Lower and Upper Bounds. *Political Analysis*, 31(4), 519–536. <https://doi.org/10.1017/pan.2022.20>
- Kubinec, R. (2023). *Ordbetareg: Ordered beta regression models with 'brms'*. <https://CRAN.R-project.org/package=ordbetareg>
- Lenth, R. V. (2025). *Emmeans: Estimated marginal means, aka least-squares means*. <https://doi.org/10.32614/CRAN.package.emmeans>
- Liu, F., & Kong, Y. (2015). zoib: An R Package for Bayesian Inference for Beta Regression and Zero/One Inflated Beta Regression. *The R Journal*, 7(2), 34. <https://doi.org/10.32614/rj-2015-019>
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Wiernik, B. M., Bacher, E., Thériault, R., & Makowski, D. (2022). *Easystats: Framework for easy statistical modeling, visualization, and reporting*. <https://easystats.github.io/easystats/>
- Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the bayesian framework. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02767>
- Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework. *Journal of Open Source Software*, 4(40), 1541. <https://doi.org/10.21105/joss.01541>
- Makowski, D., Neves, A., & Field, A. P. (2025). *Introducing the choice-confidence (CHOCO) model for bimodal data from subjective ratings: Application to the effect of attractiveness on reality beliefs about AI-generated faces*. [https://doi.org/10.31234/osf.io/z68v3\\_v1](https://doi.org/10.31234/osf.io/z68v3_v1)
- Marsman, M., & Wagenmakers, E.-J. (2016). Three Insights from a Bayesian Interpretation of the One-Sided P Value. *Educational and Psychological Measurement*. <https://doi.org/10.1177/0013164416669201>
- Martin, K., Cornero, F. M., Clayton, N. S., Adam, O., Obin, N., & Dufour, V. (2024). Vocal complexity in a socially complex corvid: Gradation, diversity and lack of common call repertoire in male rooks. *Royal Society Open Science*, 11(1), 231713. <https://doi.org/10.1098/rsos.231713>
- McCabe, C. J., Halvorson, M. A., King, K. M., Cao, X., & Kim, D. S. (2021). Interpreting Interaction Effects in Generalized Linear Models of Nonlinear Probabilities and Counts. *Multivariate Behavioral Research*, 57(2-3), 243–263. <https://doi.org/10.1080/00273171.2020.1868966>
- McElreath, R. (2020). *Statistical rethinking: A bayesian course with examples in r and STAN* (2nd ed.). Chapman; Hall/CRC. <https://doi.org/10.1201/9780429029608>
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370–384. <https://doi.org/10.2307/2344614>
- Nouvian, M., Foster, J. J., & Weidenmüller, A. (2023). Glyphosate impairs aversive learning in bumblebees. *Science of The Total Environment*, 898, 165527. <https://www.sciencedirect.com/science/article/pii/S0048969723041505>
- Paolino, P. (2001). Maximum Likelihood Estimation of Models with Beta-Distributed Dependent Variables. *Political Analysis*, 9(4), 325–346. <https://doi.org/10.1093/oxfordjournals.pan.a004873>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

- 775 R. A. Rigby, & D. M. Stasinopoulos. (2005). *Generalized additive models for location, scale and shape,(with  
776 discussion)*. 54, 507–554.
- 777 Rodrigues, B., & Baumann, P. (2025). *Rix: Reproducible data science environments with 'nix'*. [https://  
778 docs.ropensci.org/rix/](https://docs.ropensci.org/rix/)
- 779 Rohrer, J. M., & Arel-Bundock, V. (2025). *Models as prediction machines: How to convert confusing coef-  
780 ficients into clear quantities*. [https://doi.org/10.31234/osf.io/g4s2a\\_v1](https://doi.org/10.31234/osf.io/g4s2a_v1)
- 781 Shrestha, S., Sigdel, K., Pokharel, M., & Columbus, S. (2024). Big five traits predict between- and within-  
782 person variation in loneliness. *European Journal of Personality*, 08902070241239834. [https://doi.org/10.1177/08902070241239834](https://doi.org/10.<br/>783 1177/08902070241239834)
- 784 Sladekova, M., & Field, A. P. (2024). *In search of unicorns: Assessing statistical assumptions in real psy-  
785 chology datasets*. <https://doi.org/10.31234/osf.io/4rznt>
- 786 Smith, K. E., Panlilio, L. V., Feldman, J. D., Grundmann, O., Dunn, K. E., McCurdy, C. R., Garcia-  
787 Romeu, A., & Epstein, D. H. (2024). Ecological momentary assessment of self-reported kratom use,  
788 effects, and motivations among US adults. *JAMA Network Open*, 7(1), e2353401. [https://doi.org/10.1001/jamanetworkopen.2023.53401](https://doi.org/10.<br/>789 1001/jamanetworkopen.2023.53401)
- 790 Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with  
791 beta-distributed dependent variables. *Psychological Methods*, 11(1), 54–71. [https://doi.org/10.1037/1082-989x.11.1.54](https://doi.org/10.1037/<br/>792 1082-989x.11.1.54)
- 793 Team, S. D. (2023). *Stan: A probabilistic programming language*. <https://mc-stan.org>
- 794 Toftness, A. R., Carpenter, S. K., Geller, J., Lauber, S., Johnson, M., & Armstrong, P. I. (2017). Instructor  
795 fluency leads to higher confidence in learning, but not better learning. *Metacognition and Learning*,  
796 13(1), 1–14. <https://doi.org/10.1007/s11409-017-9175-0>
- 797 Vuorre, M. (2019, February 18). *How to Analyze Visual Analog (Slider) Scale Data?* [https://vuorre.com/  
posts/2019-02-18-analyze-analog-scale-ratings-with-zero-one-inflated-beta-models](https://vuorre.com/<br/>798 posts/2019-02-18-analyze-analog-scale-ratings-with-zero-one-inflated-beta-models)
- 799 Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *Rfor Data Science: Import, Tidy, Transform,  
800 Visualize, and Model Data*. O'Reilly. <https://r4ds.hadley.nz/>
- 801 Wilford, M. M., Kurpad, N., Platt, M., & Weinstein-Jones, Y. (2020). Lecturer fluency can impact students'  
802 judgments of learning and actual learning performance. *Applied Cognitive Psychology*, 34(6), 1444–  
803 1456. <https://doi.org/10.1002/acp.3724>
- 804 Wilkes, L. N., Barner, A. K., Keyes, A. A., Morton, D., Byrnes, J. E. K., & Dee, L. E. (2024). Quantifying  
805 co-extinctions and ecosystem service vulnerability in coastal ecosystems experiencing climate warming.  
806 *Global Change Biology*, 30(7), e17422. <https://doi.org/10.1111/gcb.17422>
- 807 Witherby, A. E., & Carpenter, S. K. (2022). The impact of lecture fluency and technology fluency on students'  
808 online learning and evaluations of instructors. *Journal of Applied Research in Memory and Cognition*,  
809 11(4), 500–509. <https://doi.org/10.1037/mac0000003>
- 810 Ye, M., & Heiss, A. (2023). Enforcing Boundaries: China's Overseas NGO Law and Operational Con-  
811 straints for Global Civil Society. *Working Paper*. [https://stats.andrewheiss.com/compassionate-clam/  
manuscript/output/manuscript.html](https://stats.andrewheiss.com/compassionate-clam/<br/>812 manuscript/output/manuscript.html)